The Developmental Testbed Center's Report on the inter-comparison of AFWA Operational configurations using WRFv3.3.1 and WRFv3.4

31 January 2013

Executive Summary

The Weather Research and Forecasting (WRF) model is a mesoscale numerical weather prediction system utilized in both research and operational forecasting applications. The model is configurable to the users' requirements and suitable for a broad spectrum of weather regimes. Due to the flexibility offered by the model, it is necessary to rigorously test and evaluate the performance of the model as improvements and additions to the model are added then released to the community of users. To assess the performance of WRF as it progresses from version 3.3.1 to version 3.4, the Developmental Testbed Center (DTC) performed testing and evaluation with the Advanced Research WRF (ARW) dynamic core for the two versions of the WRF system at the request of the sponsor, the Air Force Weather Agency (AFWA). The test was conducted in a functionally similar operational environment to AFWA operations, where each configuration was initialized with output from AFWA's Land Information System (LIS) as well as a 6-hour "warm start" spin-up, including the WRF Data Assimilation (WRFDA) component. The only difference between the two model configurations in the extensive testing was the WRF version number; this allows for a direct assessment of how performance changes with increasing WRF version. Version 3.3.1 of the WRF system was used as the baseline in this testing. This report focuses on the pair-wise differences between the standard verification metrics for the two configurations, including an assessment of the statistical significance (SS) and practical significance (PS). Bias-corrected root-mean-square-error (BCRMSE) and bias were evaluated for surface and upper air temperature, dew point temperature, and wind speed; Gilbert Skill Score (GSS) and frequency bias were evaluated for 3-hourly and daily quantitative precipitation forecasts (QPF). The following points summarize the SS and PS differences seen in the verification results between WRFv3.3.1 and WRFv3.4.

- Upper air temperature
  - **BCRMSE**: SS pair-wise differences mostly occur at and above 500 hPa, and favor the WRFv3.3.1 configuration; only a couple SS pair-wise differences, at 850 and 700 hPa, favor the WRFv3.4 configuration; no PS differences are observed
  - **Bias**: WRFv3.4 is SS favored at 500-200 hPa and PS favored at 100 hPa; WRFv3.3.1 is SS favored at 850-700 hPa, and frequently PS favored at 150 hPa;
- Upper air dew point temperature
  - **BCRMSE**: All SS pair-wise differences favor the WRFv3.3.1 configuration; only few of them are PS
  - **Bias**: All SS pair-wise differences favor the WRFv3.4 configuration; only one is PS
- Upper air wind speed
  - **BCRMSE**: All SS pair-wise differences favor the WRFv3.3.1 configuration; no PS pair-wise differences are observed
  - **Bias**: WRFv3.4 is generally SS favored at and below 200 hPa; WRFv3.3.1 is SS favored at and above 150 hPa; no PS differences are observed
- Surface temperature
  - **BCRMSE**: All but one SS pair-wise difference favors the WRFv3.3.1 configuration; most of the SS differences are noted for forecasts valid between 03-15 UTC; no PS differences are observed
  - **Bias**: The differences between the two configurations vary diurnally, with larger values between 03-15 UTC; all SS and PS pair-wise differences favor the WRFv3.3.1 configuration; PS differences are observed for the spring and summer aggregations
- Surface dew point temperature
  - **BCRMSE**: All but one SS pair-wise differences favor the WRFv3.3.1 configuration; none of the SS differences are PS

        - **Bias**: A majority of the SS pair-wise differences favor the WRFv3.3.1 configuration, with a few of them for the spring aggregation being PS; the WRFv3.4 configuration is SS favored at some lead times for the winter aggregation
- Surface wind speed
    - **BCRMSE**: A number of SS pair-wise differences are present; all but one favor the WRFv3.3.1 configuration, none being PS
    - **Bias**: SS pair-wise differences are mostly found between the valid times of 03-15 UTC; all but one SS difference favors the WRFv3.4 configuration; no PS is noted
- Three-hourly QPF
    - **GSS**: Very few SS pair-wise differences are noted with a majority favoring the WRFv3.3.1 configuration; the few SS differences that favor the WRFv3.4 configuration are seen for higher precipitation threshold (>1.0")
    - **Frequency Bias**: no SS pair-wise differences noted
- Daily QPF
    - **GSS**: When SS pair-wise differences occur, they tend to favor the WRFv3.3.1 configuration; the few SS pair-wise differences that favor WRFv3.4 are seen for thresholds higher than 1.25".
    - **Frequency Bias**: no SS pair-wise differences noted
- GO Index
    - The median values indicate the WRFv3.3.1 configuration is more skillful during summer, and WRFv3.4 is more skillful during winter. The two configurations are comparable for the annual, spring and fall aggregations. A number of outlier cases are also noted, most of which indicated that WRFv3.3.1 performed better than WRFv3.4.

1. Introduction

The Weather Research and Forecasting (WRF) model is a mesoscale numerical weather prediction system utilized in both research and operational forecasting applications. The model is configurable to the users' requirements and suitable for a broad spectrum of weather regimes. Due to the flexibility offered by the model, it is necessary to rigorously test and evaluate the performance of the model as improvements and additions are made and then released to the community of users. The Air Force Weather Agency (AFWA) is interested in improvements to their operational configuration (OC) as well as ensuring that changing to a newer version of WRF is not degrading the performance of their forecasts. To assess the performance of WRF as it progresses with each release, the Developmental Testbed Center (DTC) performed testing and evaluation with the Advanced Research WRF (ARW) dynamic core (Skamarock et al. 2008) for two versions of the WRF system, v3.3.1 and v3.4 in 2012. The test was conducted in a functionally similar operational environment to AFWA operations, where each configuration was initialized with output from AFWA's Land Information System (LIS) as well as a 6-hour "warm start" spin-up, including data assimilation. The only difference between the two model configurations in the extensive testing was the WRF version number. This allows for a direct assessment of how performance changes with increasing WRF version; version 3.3.1 of the WRF system was used as the baseline in this testing. To assess the performance of these configurations, verification statistics were computed for the two configurations, and the analysis was based on the objective statistics of the model output. In addition to documenting the performance of the two configurations against each other, both of these configurations have been designated as DTC Reference Configurations (RCs) and the results are available to the WRF community through the DTC RC webpage.

2. Experiment Design

For this test, the end-to-end forecast system consisted of the WRF Preprocessing System (WPS), WRF Data Assimilation (WRFDA) system, WRF, Unified Postprocessor (UPP) and the NCAR Command Language (NCL) for graphics generation. Post-processed forecasts were verified using the Model Evaluation Tools (MET). In addition, the full data set was archived and made available for dissemination to the user community. The codes utilized were based on the official released versions of WPS (v3.3.1,

v3.4), WRFDA (v3.3.1, v3.4), WRF (v3.3.1, v3.4), UPP (v1.1), and MET (v4.0). MET included relevant bug fixes that were checked into the code repository prior to testing.

2.1 Forecast Periods

Forecasts were initialized every 36 hours from 1 July 2011 through 29 June 2012, consequently creating a default of initialization times including both 00 and 12 UTC, for a total of 244 cases (see Appendix A for a list of the cases). The forecasts were run out to 48 hours with output files generated every 3 hours.

The tables below list the forecast initializations that failed to complete the end-to-end process; the missing data and reason for failure is described in the table. All missing forecasts were due to missing or bad input data sets, not model crashes. A total of 239 cases ran to completion and were used in the verification results.

Missing forecasts:

Affected Cycle	Missing data	Reason
2011080112	WRF output	Bad SST input data
2011082400	WRF output	Missing SST input data
2012050312	WRF output	Missing GFS input data
2012050612	WRF output	Bad obs_gts input data
2012060400	WRF output	Bad SST input data

Missing verification:

Affected Cycle	Missing data	Reason
2011072500	Missing 3-h QPF verification for 18 – 21-h Missing 24-h QPF verification for 36-h	Missing ST2 analysis

2.2 Initial and Boundary Conditions

Initial conditions (ICs) and lateral boundary conditions (LBCs) were derived from the 0.5° x 0.5° Global Forecast System (GFS). Output from AFWA's LIS running with version 2.7.1 of the Noah LSM, was used to initialize the lower boundary conditions (LoBCs). In addition, a daily, real-time sea surface temperature product from Fleet Numerical Meteorology and Oceanography Center (FNMOC) was used to initialize the sea surface temperature (SST) field for the forecasts.

The time-invariant components of the LoBCs (topography, soil, vegetation type, etc.) were derived from United States Geological Survey (USGS) input data and were generated through the *geogrid* program of WPS. The *avg_tsfc* program of WPS was also used to compute the mean surface air temperature in order to provide improved water temperature initialization for lakes and smaller bodies of water in the domain that are further away from an ocean.

A 6-hour "warm start" spin-up procedure (Fig. 1) preceded each forecast. Data assimilation using WRFDA was conducted at the beginning and end of the 6-hour window using observation data files provided by AFWA. At the beginning of the data assimilation window, the GFS derived initial conditions were used as the model background, and at the end of the window, the 6-hour WRF forecast initialized by the WRFDA analysis was used. After each WRFDA run, the LBCs initially derived from GFS were updated and used in the subsequent forecasts.

Seasonal, domain-specific model background error statistics (BE) files were created and used in WRFDA. To create the appropriate BE files, cold-start WRF forecasts were conducted on the 15 km grid twice daily for 15 days each season. Essentially, this was 30 forecasts per season, or 120 total forecasts (24-h forecasts, in 12-h increments). The *gen_be* utility in WRFDA was then used to generate BE files from those model runs.

2.3 Model Configuration Specifics

2.3.1 Domain Configuration

A 15-km contiguous U.S. (CONUS) grid was employed for this test. The domain (Fig. 2) was selected such that it covers complex terrain, plains, and coastal regions spanning from the Gulf of Mexico, north, to Central Canada in order to capture diverse regional effects for worldwide comparability. The domain has 403 x 302 gridpoints, for a total of 121,706 gridpoints. The Lambert-Conformal map projection was used and the model was configured to have 56 vertical levels (57 sigma entries) with the model top at 10 hPa.

2.3.2 Other Aspects of Model Configuration

The table below lists AFWA's current OC that was used in this testing. The model configuration based on version 3.3.1 of the WRF system will be referred to as WRFv3.3.1, while the companion configuration will be referred to as WRFv3.4.

	Current AFWA OC (AFWA)
Microphysics	WRF Single-Moment 5 scheme
Radiation LW and SW	RRTM/Dudhia schemes
Surface Layer	Monin-Obukhov similarity theory
Land-Surface Model	Noah
Planetary Boundary Layer	Yonsei University scheme
Convection	Kain-Fritsch scheme

Both configurations were run with a long timestep of 90 s, and an acoustic step of 4 was used. Calls to the boundary layer, and microphysics were performed every time step, whereas the cumulus parameterization was called every 5 minutes and every 30 minutes for the radiation.

The ARW solver offers a number of run-time options for the numerics, as well as various filter and damping options (Skamarock et al. 2008). The ARW was configured to use the following numeric options: 3rd-order Runge-Kutta time integration, 5th-order horizontal momentum and scalar advection, and 3rd-order vertical momentum and scalar advection. In addition, the following filter/damping options were utilized: three-dimensional divergence damping (coefficient 0.1), external mode filter (coefficient 0.01), off-center integration of vertical momentum and geopotential equations (coefficient 0.1), vertical-velocity damping, and a 5-km-deep diffusive damping layer at the top of the domain (coefficient 0.02). Positive-definite moisture advection was also turned on.

In the extensive testing, it was discovered that two namelist options in WRFv3.4 are not compatible [hypsoemtric_opt set to 2 (default) and adjust_heights set to true (default is false)]. The hypsoemtric_opt = 2 option is new to WRF as of version 3.4 and was incorporated as the default. The AFWA OC that the DTC has been testing over several years (i.e., versions prior to v3.4) has the adjust_heights namelist option set to true. The original set up for this test caused several model crashes in the summer months of the year-long test due to the incompatibility of the namelist options. A new check to assure adjust_heights set to true is not used with hypsoemtric_opt set to 2 has been added to the WRF code repository due to discovering this issue and will be released with the next official code distribution. Users will want to examine their v3.4 namelists to ensure they do not use these options concurrently.

Appendix B provides relevant portions of the *namelist.input* file.

2.4 Post-processing

The *unipost* program within UPP was used to destagger the forecasts, to generate derived meteorological variables, and to vertically interpolate fields to isobaric levels. The post-processed files included two- and three-dimensional fields on constant pressure levels, both of which were required by the plotting and

verification programs. Three-dimensional post-processed fields on model native vertical coordinates were also output and used to generate graphical forecast sounding plots.

3. Model Verification

The MET package was used to generate objective model verification. MET is comprised of grid-to-point verification, which was utilized to compare gridded surface and upper-air model data to point observations, as well as grid-to-grid verification, which was utilized to verify QPF. Verification statistics generated by MET for each retrospective case were loaded into a MySQL database. Data was then retrieved from this database to compute and plot specified aggregated statistics using routines developed by the DTC in the statistical programming language, R.

Several domains were verified for the surface and upper air, as well as precipitation variables. Area-average results were computed for the CONUS domain, as well as the 14 sub-domains shown in Fig. 3. Only the CONUS domain is described in detail for this report, with some brief discussions on the sub-domain results of surface verification. However, all CONUS, East, West and sub-domain results are available on the DTC website (http://www.dtcenter.org/config/v3.4/ARW_PS4.1.1.1.2.1.1_LIS2.7.1/). In addition to the regional stratification, the verification statistics were also stratified by vertical level and lead time for the 00 UTC and 12 UTC initialization hours combined, and by forecast lead time and precipitation threshold for 00 UTC and 12 UTC initialized forecasts individually for surface fields in order to preserve the diurnal signal.

Each type of verification metric is accompanied by confidence intervals (CIs), at the 99% level, computed using the appropriate statistical method. Both configurations were run for the same cases allowing for a pair-wise difference methodology to be applied, as appropriate. The CIs on the pair-wise differences between statistics for the two configurations objectively determines whether the differences are statistically significant (SS); if the CIs on the median pair-wise difference statistics include zero, the differences are not statistically significant. Due to the nonlinear attributes of frequency bias, it is not amenable to a pair-wise difference comparison. Therefore, the more powerful method to establish SS could not be used and, thus, a more conservative estimate of SS was employed based solely on whether the aggregate statistics, with the accompanying CIs, overlapped between the two configurations. If no overlap was noted for a particular threshold, the differences between the two configurations were considered SS.

Due to the large number of cases used in this test, many SS pair-wise differences were anticipated. In many cases, the magnitude of the SS differences was quite small and did not yield practically meaningful results. Therefore, in addition to determining SS, the concept of establishing practical significance (PS) was also utilized for this test. PS was determined by filtering results to highlight pair-wise differences greater than the operational measurement uncertainty requirements and instrument performance as specified by the World Meteorological Organization (WMO; http://www.wmo.int/pages/prog/www/IMOP/publications/CIMO-Guide/1st-Suppl-to-7th_draft/pdf/Annex_I_1B.pdf). To establish PS between the two configurations, the following criteria was applied: temperature and dew point temperature differences greater than 0.1 K and wind speed differences greater than 0.5 m s^{-1}. PS was not considered for metrics used in precipitation verification [i.e., Gilbert Skill Score (GSS) or frequency bias] because those metrics are calculated via a contingency table, which is based on counts of yes and no forecasts.

3.1 Temperature, Dew Point Temperature, and Winds

Forecasts of surface and upper air temperature, dew point temperature, and wind were bilinearly interpolated to the location of the observations (METARs and RAOBS) within the National Centers for Environmental Prediction (NCEP) North American Data Assimilation System (NDAS) prepbufr files. Objective model verification statistics were then generated for surface (using METAR) and upper air (using RAOBS) temperature, dew point temperature, and wind. Because shelter-level variables are not available from the model at the initial time, surface verification results start at the 3-hour lead time and go

out 48 hours by 3-hour increments. For upper air, verification statistics were computed at the mandatory levels using radiosonde observations and computed at 12-hour intervals out to 48 hours. Because of known errors associated with radiosonde moisture measurements at high altitudes, the analysis of the upper air dew point temperature verification focuses on levels at and below 500 hPa. Bias and bias-corrected root-mean-square-error (BCRMSE) were computed separately for surface and upper air observations. The CIs were computed from the standard error estimates about the median value of the stratified results using a parametric method and a correction for first-order autocorrelation.

3.2 Precipitation

For the QPF verification, a grid-to-grid comparison was made by first bilinearly interpolating the precipitation analyses to the 15-km model integration domain. This regridded analysis was then used to evaluate the forecast. Accumulation periods of 3 and 24 hours were examined. NCEP Stage II analysis was used as the observational dataset, and the data is available in hourly, 6-hourly, and 24-hourly accumulations. For this test, hourly data was summed for the 3-hour QPF verification, and daily QPF verification utilized the 24-hour accumulation files. The 24-hour accumulation observations are valid at 12 UTC; therefore, the daily QPF was examined for the 24- and 48-hour lead times for the 12 UTC initializations and 36-hour lead time for the 00 UTC initializations. Traditional verification metrics computed included the GSS and frequency bias. For the precipitation statistics, a bootstrapping CI method was applied.

3.3 GO Index

Skill scores (S) were computed for wind speed (at 250 hPa, 400 hPa, 850 hPa and surface), dew point temperature (at 400 hPa, 700 hPa, 850 hPa and surface), temperature (at 400 hPa and surface), height (at 400 hPa), and mean sea level pressure, using root-mean-square-error (RMSE) for both the WRFv3.3.1 and WRFv3.4 configurations using the formula:

$$S = 1 - \frac{(RMSE_{v3.4})^2}{(RMSE_{v3.3.1})^2}$$

For each variable, level, and forecast hour, predefined weights (w_i), shown in the table below, were then applied and a weighted sum, S_W, was computed

Variable	Level	Weights by lead time			
		12 h	24 h	36 h	48 h
Wind Speed	250 hPa	4	3	2	1
	400 hPa	4	3	2	1
	850 hPa	4	3	2	1
	Surface	8	6	4	2
Dew Point Temperature	400 hPa	8	6	4	2
	700 hPa	8	6	4	2
	850 hPa	8	6	4	2
	Surface	8	6	4	2
Temperature	400 hPa	4	3	2	1
	Surface	8	6	4	2
Height	400 hPa	4	3	2	1
Pressure	Mean sea level	8	6	4	2

where,

$$S_w = \frac{1}{\sum_i w_i}\left(\sum_i (w_i S_i)\right)$$

Once the weighted sum of the skill scores, S_w, was computed, the Index value (N) is defined as:

$$N = \sqrt{\frac{1}{1 - S_w}}$$

Given this definition, which is based on the General Operations (GO) Index, values (N) less than one indicate the WRFv3.3.1 configuration has higher skill and values greater than one indicate the WRFv3.4 configuration has higher skill.

4. Verification Results

Differences between the two configurations are computed by subtracting WRFv3.3.1 from WRFv3.4. BCRMSE is always a positive quantity and a perfect score is zero. As a result, differences that are negative (positive) indicate the WRFv3.4 (WRFv3.3.1) configuration has a lower BCRMSE and is favored. Bias also has a perfect score of zero but can have positive or negative values; therefore, when looking at the pair-wise differences it is important to also note the magnitude of the bias in relation to the perfect score for each individual configuration to know which configuration has a smaller bias and is, thus, favored. In additional to time series plots, the surface verification statistics are also available on regional plots. While only the regional distribution of bias is shown in figures in this report, the regional distribution of BCRMSE is also briefly discussed. For GSS, the perfect score is one and the no-skill score is zero. Thus, if the pair-wise difference is positive (negative), the WRFv3.4 (WRFv3.3.1) configuration has a higher GSS and is favored. For the frequency bias of QPF, a perfect score is one. A breakdown of the configuration with SS and PS better performance by variable, season, statistics metric, initialization hour, forecast lead time, and pressure level is summarized in Tables 1-8, where the favored configuration is highlighted. All verification plots generated (by plot type, metric, lead time, threshold, season, etc.) can be viewed on the DTC webpage.

4.1 Upper Air

4.1.1 Temperature BCRMSE and bias

Both the WRFv3.4 and WRFv3.3.1 configurations, regardless of forecast lead time or temporal aggregation, have a minimum in BCRMSE values between 500 and 300 hPa with the largest errors at both the lower and upper levels (Fig. 1). For all vertical levels, there is an error growth with lead time. SS pair-wise differences between the two configurations mostly occur at and above 500 hPa, and favor the WRFv3.3.1 configuration (see Table 1). The summer aggregation displays the most SS pair-wise differences while the winter and spring aggregations display the fewest. Only a couple SS pair-wise differences, at 850 hPa and 700 hPa, favor the WRFv3.4 configuration. None of the SS pair-wise differences for temperature BCRMSE are PS.

For all temporal aggregations and forecast lead times, both configurations have a cold bias at 850 and 700 hPa, which transitions to a warm bias with height (Fig. 5). The transition to a warm bias occurs around 500 - 400 hPa, depending on the configuration, season, and forecast lead time. The bias at initialization time (LT=0 h) reflects the error at the end of the 6 h data assimilation window. There appears to be a large variability in the bias at 150 hPa, which exhibits anything from no bias (CIs encompass zero) to a warm bias, or a cold bias, depending on the configuration, forecast lead time and seasonal aggregations. Meanwhile, regardless of configuration, lead time, and seasonal aggregation, there is a vertical local minimum (larger cold bias or smaller warm bias) at 150 hPa. In general, the WRFv3.4 forecast temperatures appear colder than those of WRFv3.3.1. As a result, at 850 – 700 hPa, the WRFv3.4 configuration has a larger cold bias than the WRFv3.3.1 configuration, with some SS but no PS differences (see Table 1). At the mid- to upper-levels except for 150 hPa, the WRFv3.4 configuration has a smaller warm bias than WRFv3.3.1, with all pair-wise differences at 100 hPa being PS. At 150 hPa, the WRFv3.3.1 configuration is PS favored for all temporal aggregations except summer. For summer, the WRFv3.4 configuration is SS or PS favored for all vertical levels at and above 500 hPa.

4.1.2 Dew Point Temperature BCRMSE and bias

The dew point temperature BCRMSE increases as pressure decreases and as lead time increases, for both configurations and the annual, spring and summer aggregations (Fig. 3). For the fall and winter aggregations, a slight decrease in the median BCRMSE value is noted as pressure decreases from 700 hPa to 500 hPa (not shown). Few or no SS pair-wise differences are noted in each seasonal aggregation, all showing WRFv3.3.1 as the favored configuration. A couple of the SS differences are PS (see Table 2).

At forecast initialization time, a high (wet) bias is observed for both configurations at all vertical levels for all temporal aggregations (Fig. 7). The bias at 500 hPa increases with forecast lead time, but the bias at both 850 hPa and 700 hPa decreases with lead time for most temporal aggregations. While the bias at 500 hPa remains as a high bias for all lead time and temporal aggregations, the bias at 700 hPa and 850 hPa becomes neutral (CIs encompass zero) or a low bias at some lead time and seasonal aggregations. For the summer aggregation, at 700 hPa, the 12-24 h forecasts are unbiased and the 36-48 h forecasts have a cold bias; at 850 hPa, the forecasts are unbiased through the 12 – 48 h lead times. All SS pair-wise differences favor the WRFv3.4 configuration. Only one PS pair-wise difference is noted, which occurred in the fall aggregation, at 850 hPa and 48 h lead time.

4.1.3 Wind Speed BCRMSE and bias

The vertical distribution of wind speed BCRMSE for all temporal aggregations and lead times are very similar for both configurations. The lowest values of wind speed BCRMSE are typically seen at the 850 or 700 hPa level with values increasing to a maximum around 300 hPa (fall and winter) or 200 hPa (spring and summer) level before decreasing again further aloft (Fig. 8). The BCRMSE values grow with lead time for all temporal aggregations. For the same lead time, the BCRMSE values are generally larger for the fall/winter than for the spring/summer (not shown). A number of SS pair-wise differences are noted, all favoring the WRFv3.3.1 configuration; however, the magnitudes of the differences are small, with none being PS (see Table 3).

The bias is mostly negative (i.e. winds are too light) at initialization with the magnitude largest at 850 hPa and decreasing from 500 hPa to 150 hPa (Fig. 9). There is a significant change in the vertical distribution of bias from the initialization to the 12 h forecast. As forecast lead time further increases, the bias at lower levels (below 400 hPa) decreases in magnitude (becoming less negative). For several lead times, both configurations have unbiased forecasts (i.e., the CIs encompass zero) at 850 hPa for all temporal aggregations. As pressure decreases, the bias generally becomes more negative up to 200 hPa, where the largest magnitude of bias is noted. Above 200 hPa, the bias becomes less negative with height, and for certain forecast lead times, the forecasts become unbiased or display a high bias. In general, the WRFv3.4 configuration tends to produce higher wind than WRFv3.3.1, and as a result, it has a smaller bias in the negative bias regime (lower to mid vertical levels) and larger bias in the positive bias regime (highest vertical levels). There are many SS pair-wise differences, favoring the WRFv3.4 configuration at lower to mid levels and the WRFv3.3.1 configuration at and above 150 hPa; however, again there are no PS pair-wise differences (see Table 3).

4.2 Surface

4.2.1 Temperature BCRMSE and bias

The surface temperature BCRMSE displays a general increase with lead time for both the 00 and 12 UTC initializations and for all seasonal aggregations (Fig. 10). In addition, a diurnal signal is noted, with the strength of the signal dependent on the season. For the annual aggregation, the lowest BCRMSE values occur at times valid around midnight (06 – 09 UTC), while the largest error values are seen at times valid in the early morning (15 UTC). The differences between the two configurations also exhibit diurnal variations, with larger differences between 03-15 UTC valid times. A number of SS pair-wise differences, none being PS, are seen for both initializations dependent on temporal aggregation and lead time (see Table 4), with all but one favoring the WRFv3.3.1 configuration.

In general, a cold bias in surface temperature is observed at a majority of forecast lead times for both 00 and 12 UTC initializations (Fig. 11). With increasing forecast lead time, the bias displays a strong diurnal modulation on top of a gentler increasing trend. For both configurations, both initializations, and all temporal aggregations, the often observed cold bias magnitude is largest at valid time near 18 – 00 UTC (i.e., during the afternoon) and smallest at 06-12 UTC (night). This indicates that both configurations under-predict the diurnal variation of the surface temperature, in addition to under-predicting the surface temperature at given times. The cold bias in summer is more pronounced on average but exhibits less diurnal modulation, whereas in winter the cold bias is significantly smaller during the overnight hours. The differences between the two model configurations also vary diurnally, with larger differences between 03-15 UTC. SS pair-wise differences are found for all temporal aggregations and at most lead times, all favoring the WRFv3.3.1 configuration. Many of the SS pair-wise differences are PS, mostly for the spring aggregation.

Sub-domain verification of surface temperature shows a cold bias for almost all regions and most lead times (Fig. 12), with the only exception of a neutral or small warm bias at night for some western regions. The biases for all sub-domains exhibit a diurnal cycle, with the largest bias at 10 – 16 local time and smallest at 01 - 07 local time. At 00 UTC validation time, the largest magnitude of bias is noted in GRB and the mountainous regions, smaller magnitudes are found in the Great Plains and coastal regions. At 12 UTC validation time, the western regions have a smaller cold bias than the eastern regions; some western regions display a warm bias for certain temporal aggregations (such as NPL in the annual, fall and winter aggregations, GRB in summer and fall). The problem of under-predicting the diurnal extend of surface temperature is most severe for GRB where large cold bias is noticed during the afternoon and small or neutral bias is seen before sunrise. The WRFv3.4 configuration tends to produce a more pronounced cold bias in most regions (except NPL) at most lead time, and therefore is less favored.

4.2.2 Dew Point Temperature BCRMSE and bias

Similar to surface temperature BCRMSE, an increase with lead time in dew point temperature BCRMSE is noted, for both initializations and all temporal aggregations (Fig. 13); a diurnal trend exists in all temporal aggregations, but is the most pronounced in the spring aggregation and least pronounced in the fall aggregation (not shown). The differences between the two configurations are generally small, leading to a number of SS but no PS pair-wise differences (see Table 5). All but one of the SS pair-wise differences favor the WRFv3.3.1 configuration.

The sign of the dew point temperature bias is sensitive to temporal aggregation and forecast lead time for both configurations. A diurnal cycle is noted in the bias for both configurations, both initializations and all temporal aggregations, with the spring and summer aggregations exhibiting the largest amplitudes and winter aggregation exhibiting the smallest (Fig. 14). For the annual aggregation, both configurations generally have high biases for forecasts valid 18 – 00 UTC and low biases for forecasts valid 06 – 12 UTC. Thus, the diurnal cycle of the dew point temperature bias is out of phase with that of the temperature bias. For most of the temporal aggregations, there is also a trend of the bias drifting toward the low side with increasing forecast lead time. The differences between the two configurations do not seem to increase with lead time, though. A majority of the SS pair-wise differences favors the WRFv3.3.1 configuration (see Table 5), with some of the differences being PS (for the spring aggregation). The WRFv3.4 configuration is SS favored at some lead times for the winter aggregation.

For the sub-domains, larger values of dew point temperature BCRMSE are usually seen in the western regions, especially SWD and SMT (not shown). The regional bias distribution shows that, at 00 UTC valid time (wet bias regime on average) the largest wet bias is seen in GRB and SMT (Fig. 15), which happen to be the regions of largest cold temperature bias for that time. At the 12 UTC valid time (dry bias regime on average), the largest dry bias is found at the west coast regions, SWC and NWC. The WRFv3.4 configuration appears to be wetter (e.g. in SWD) at 00 UTC and drier (e.g. in SWC) at 12 UTC, giving slightly larger bias values, in general.

4.2.3 Wind BCRMSE and bias

For both configurations and all temporal aggregations, the surface wind speed BCRMSE increases with forecast lead time (Fig. 16). A diurnal signal is clearly noted in the spring and summer aggregations, but less so in the fall and winter aggregations. For the spring, summer and annual aggregations, both 00 and 12 UTC initializations, both configurations have the smallest wind speed BCRMSE around 12 UTC validation time, while the largest errors occur around 00 UTC. The differences between the two configurations are small in magnitude; a number of SS pair-wise differences are present, all but one favoring the WRFv3.3.1 configuration, but no PS differences are noted (see Table 6).

A high wind speed bias up to 2 ms^{-1} is noted for both configurations at nearly all forecast lead times, regardless of initialization time or temporal aggregation (Fig. 17). In addition, a prominent diurnal trend is noted for both configurations and all initialization times and temporal aggregations. A larger high bias is typically seen in the forecasts for the overnight hours (03-12 UTC) as compared to the daytime hours (15-21 UTC). Apart from the diurnal trend, there is a drift of the bias toward higher wind with increasing lead time. For both initializations (00 UTC shown in Fig. 17), the bias for the winter aggregation appears to be larger than that for the summer. The WRFv3.4 configuration tends to have smaller high biases for most of the forecast times, and as a result gives better performance. Only SS pair-wise differences are noted, mostly during evening and nighttime (larger high bias regime); all but one show that the WRFv3.4 configuration is favored (see Table 6). No PS pair-wise differences are noted.

The sub-domain wind BCRMSE (not shown) in the west and MDW are generally larger than those in the east; the largest BCRMSE is found in SMT. The largest wind biases are found for 12 UTC valid times in the western regions of SMT, GRB and NWC, and for 00 UTC valid times in the Midwest and eastern regions of MDW, APL, SEC and NEC (Fig. 18). It is worth noting that at 00 UTC, the East Coast is already in the nighttime regime, and at 12 UTC the West Coast is still in the nighttime regime (highest bias are observed during overnight hours). The lowest overall wind bias is found at SPL. The WRFv3.4 configuration generally gives smaller wind biases than WRFv3.3.1 in the Midwest and east.

4.2.4 3-hourly QPF GSS and bias

For both configurations, both initializations and all forecast lead times, the 3-hourly QPF GSS steadily decreases as the threshold increases from 0.01″ to 1.0″ (Fig. 19). Similar behavior is seen in the base rate, which is a measure of the observed grid box events to the total number of grid boxes in the domain. Regardless of season, higher base rates are seen at the lower thresholds, and lower base rates are seen at the higher thresholds, due the infrequency of high-precipitation events. Therefore, CIs are often larger for both GSS and bias at high thresholds. In general, the highest base rates of events are noted during the summer season during late afternoon hours, regardless of threshold. The performance differences between the two model configurations are mostly insignificant when considering 3-hourly QPF GSS, with a number of SS pair-wise differences (see Table 7). In general, most of the SS pair-wise differences favor the WRFv3.3.1 configuration. Those few SS differences that favor the WRFv3.4 configuration are all for higher precipitation threshold (>1.0”).

A SS high bias in 3-hourly QPF frequency, typically in the range of 1.2 - 2.0, is seen for thresholds below 0.25″, for both initializations, most forecast lead time and temporal aggregations (Fig. 20). An exception is a neutral bias across all thresholds in the forecasts valid at 12 UTC in the spring and summer aggregations (not shown). At higher thresholds (>0.25″), the bias error either continues to be a high bias or becomes neutral with the generally large CIs encompassing one. There is an indication that the bias is higher for 00 UTC than for 12 UTC validation time. There are no SS pair-wise differences in the 3-hourly QPF frequency bias between the two configurations, indicating the two configurations perform comparatively in the statistical sense.

4.2.5 Daily Precipitation GSS and bias

Similar to GSS for 3-hourly QPF, GSS for daily QPF tends to decrease with increasing threshold for both configurations, both initializations, most lead times, and all temporal aggregations (Fig. 21). The base

rate decreases with increasing threshold, with very few observed events at the highest thresholds. Therefore, the width of the CIs increases with increasing threshold. The performance differences between the two configurations generally favor WRFv3.3.1, though only a smaller number of SS pair-wise differences are noted (see Table 8) for each temporal aggregation. All but a few SS pair-wise differences favor the WRFv3.3.1 configuration. The few SS pair-wise differences that favor WRFv3.4 are for thresholds higher than 1.25".

Both configurations generally have a high bias for both initializations and all lead times, at all but the lowest (0.01") and highest (3.0") thresholds. At the lowest and highest thresholds, the bias is neutral with the CIs encompassing one for some aggregations and lead times (Fig. 22). The bias in the summer aggregation tends to be smaller than that in the winter for most thresholds. No SS pair-wise differences are found between the two configurations for frequency bias of daily QPF.

4.3 GO Index

For the annual, fall and spring aggregations and both the 00 and 12 UTC initializations, the GO Index median values are close to one, and the associated CIs, indicated by the width of the notches on the boxplot, encompass one; thus, WRFv3.4 and WRFv3.3.1 perform comparatively (Fig. 23). For both initializations, the summer aggregation displays the lowest GO Index median values, around 0.995, indicating the WRFv3.4 configuration is less skillful than the WRFv3.3.1 configuration in summer. The winter aggregation has the highest median values, around 1.003, indicating the WRFv3.4 configuration is more skillful than the WRFv3.3.1 configuration in winter. It is also noted, a number of outliers exist, especially for the annual, winter and spring aggregations. Most of the outliers have GO index values significantly lower than one, indicating that WRFv3.3.1 outperform WRFv3.4 for these particular cases.

5. Summary

In this end-to-end sensitivity test, two WRF-ARW configurations were tested and evaluated. One configuration employed AFWA's OC based on WRF version 3.3.1, and the other configuration replaced WRF version 3.3.1 with WRF version 3.4 while keeping the physics options and initialization datasets intact. The only namelist difference was that *adjust_heights*, which was set to true in version 3.3.1, was set to false in version 3.4, in order to be compatible with the default value of *hypsometric_opt* (set to 2) in version 3.4. Each configuration included a 6-hour data assimilation procedure for which the respective version of WRFDA was used. The two configurations were run over an identical set of cases spanning one year. Performances of the two configurations were evaluated and compared in attempt to assess the potential overall impacts of upgrading the WRF-ARW version for AFWA.

Pair-wise differences were computed for several verification metrics, and an assessment of SS and PS were completed. Overall, there were a significant number of SS pair-wise differences between the two configurations, but most of them were not PS. Whereas which configuration was favored may depend on verification metric, temporal aggregation, initialization time, vertical level, forecast lead time, and threshold, a majority of SS and PS pair-wise differences indicated the WRFv3.3.1 configuration out-performed the WRFv3.4 configuration. There was also an indication that the WRFv3.4 configuration did better than WRFv3.3.1 in terms of bias (except for surface temperature and dew-point), but worse in terms of BCRMSE. The overall skill represented by the GO Index showed the two configurations were comparable in performance for the annual, spring and fall aggregations; while the WRFv3.3.1 configuration was the better performer of the two configurations for the summer aggregation and WRFv3.4 was the better configuration for the winter aggregations.

6. References

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3, NCAR Tech Note, NCAR/TN–475+STR, 113 pp.

Table 1. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for upper air temperature BCRMSE and bias by pressure level, season, and forecast lead time for the 00 UTC and 12 UTC initializations combined over the CONUS verification domain.

Upper Air Temperature		Annual				Summer				Fall				Winter				Spring			
		f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	--	--	--	--	v3.4	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	700	--	--	--	--	--	--	--	--	v3.4	--	--	--	--	--	--	--	--	--	--	--
	500	v3.3.1	--	v3.3.1	v3.3.1	--	--	v3.3.1	--	--	--	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--
	400	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	300	v3.3.1	--	--	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	v3.3.1	--	--	--
	200	--	--	--	v3.3.1	v3.3.1	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--
	150	--	--	v3.3.1	v3.3.1	--	--	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	--	--	--	--	--	--	v3.3.1	--	v3.3.1
	100	v3.3.1	v3.3.1	v3.3.1	--	--	v3.3.1	v3.3.1	--	--	v3.3.1	--	--	--	--	v3.3.1	--	--	--	--	--
Bias	850	v3.3.1	v3.3.1	--	--	v3.3.1	--	--	--	--	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	--	--	--
	700	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--	--	--	--	--	v3.4
	500	v3.4	v3.4	v3.4	--	v3.4	v3.4	v3.4	v3.4	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--	--	--	v3.4	--	--
	400	v3.4	v3.4	v3.4	--	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	--	v3.4	v3.3.1	--	--	v3.4	v3.4	--	--
	300	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	--	v3.4	v3.4	v3.4	--	v3.4	v3.4	--	--
	200	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4
	150	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1 *	v3.4	v3.4	v3.4	v3.4 *	v3.3.1	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1	v3.3.1 *	v3.3.1 *	v3.3.1 *
	100	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *	v3.4 *

Table 2. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for upper air dew point temperature BCRMSE and bias by pressure level, season, and forecast lead time for the 00 UTC and 12 UTC initializations combined over the CONUS verification domain.

Upper Air Dew Point Temperature		Annual				Summer				Fall				Winter				Spring			
		f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	--	--	--	--	--	--	v3.3.1 *	--	--	--	--	--	--	--	--
	700	--	--	v3.3.1	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	v3.3.1 *
	500	--	v3.3.1	--	--	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--
Bias	850	v3.4	v3.4	v3.4	v3.4	--	--	v3.4	--	v3.4	v3.4	--	v3.4 *	v3.4	v3.4	--	--	v3.4	--	--	--
	700	v3.4	v3.4	--	--	v3.4	--	--	--	v3.4	--	--	--	--	--	--	--	--	v3.4	--	--
	500	--	--	v3.4	--	--	--	--	--	v3.4	--	v3.4	--	--	--	--	--	--	--	--	--

Table 3. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for upper air wind BCRMSE and bias by pressure level, season, and forecast lead time for the 00 UTC and 12 UTC initializations combined over the CONUS verification domain.

Upper Air Wind Speed		Annual				Summer				Fall				Winter				Spring			
		f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48	f12	f24	f36	f48
BCRMSE	850	--	v3.3.1	--	--	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	v3.3.1
	700	--	v3.3.1	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	v3.3.1	--	--	--	--	--
	500	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	--	--	--	--	v3.3.1	--	v3.3.1	--	--	v3.3.1	--	--	--	--	--
	400	--	--	v3.3.1	v3.3.1	--	--	--	v3.3.1	--	--	--	--	--	--	v3.3.1	--	--	--	--	--
	300	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	v3.3.1
	200	--	--	v3.3.1	--	--	--	v3.3.1	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--
	150	v3.3.1	--	--	v3.3.1	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--
	100	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	--	--	--	--	--	--	--	v3.3.1	--	--
Bias	850	v3.4	v3.4	--	v3.4	--	--	--	--	v3.4	v3.4	--	v3.4	--	v3.4	--	--	--	--	--	v3.3.1
	700	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	v3.4	v3.4	v3.4	v3.4	v3.4	--	v3.4	--	v3.4	--	--	--
	500	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	--	v3.4	v3.4	v3.4	--	--	v3.4	--	v3.4	v3.4	v3.4	v3.4	v3.4
	400	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	--	v3.4	--	--	--	v3.4	v3.4	--	v3.3.1	v3.4	v3.4	v3.4	v3.4
	300	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	v3.4	v3.4	v3.4	--	v3.4	v3.4	--	v3.4	v3.4	v3.4	v3.4	v3.4
	200	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4
	150	v3.4	v3.3.1	v3.3.1	v3.3.1	v3.4	v3.3.1	v3.4	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.4	v3.3.1	v3.3.1	v3.3.1	v3.4	v3.3.1	v3.3.1	v3.3.1
	100	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.4	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.4	v3.3.1	v3.3.1

Table 4. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for surface temperature BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Surface Temperature			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	v3.3.1
		Summer	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	--
		Fall	v3.3.1	--	--	--	v3.3.1	v3.3.1	--	--	v3.3.1	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Winter	--	--	--	--	v3.3.1	v3.3.1	--	--	--	--	--	--	--	v3.3.1	--	v3.3.1
		Spring	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	12 UTC Initializations	Annual	v3.3.1	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Summer	--	v3.4	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Fall	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--	v3.3.1	--	v3.3.1	--	--	--	--	--	--	--
		Winter	v3.3.1	--	--	v3.3.1	--	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	v3.3.1
		Spring	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Bias	00 UTC Initializations	Annual	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Summer	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Fall	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Winter	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Spring	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1 *	--	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1 *	--	--
	12 UTC Initializations	Annual	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Summer	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1 *
		Fall	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Winter	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Spring	v3.3.1 *	v3.3.1	--	--	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1	v3.3.1 *	v3.3.1 *	v3.3.1	v3.3.1	v3.3.1 *	v3.3.1 *	v3.3.1 *	v3.3.1 *

Table 5. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for surface dew point temperature BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Surface Dew Point Temperature			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	v3.3.1	v3.3.1	v3.3.1	--	--	v3.3.1	v3.3.1	v3.3.1	--	--	--	--	--	--	v3.3.1	v3.3.1
		Summer	--	v3.3.1	v3.3.1	--	--	--	--	--	--	--	v3.3.1	--	--	--	--	--
		Fall	--	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--
		Winter	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		Spring	--	--	--	--	--	--	--	--	--	--	--	--	--	--	v3.3.1	v3.3.1
	12 UTC Initializations	Annual	v3.3.1	v3.3.1	--	v3.3.1	--	--	--	--	--	--	v3.3.1	v3.3.1	--	--	--	--
		Summer	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	--	v3.3.1	--	--	--	v3.3.1	v3.3.1	--	v3.3.1	--	v3.3.1
		Fall	v3.3.1	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	v3.4	--	--
		Winter	v3.3.1	--	--	--	--	--	--	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--	--	v3.3.1	--
		Spring	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Bias	00 UTC Initializations	Annual	--	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.4	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--
		Summer	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--
		Fall	--	--	v3.3.1	v3.3.1	v3.3.1	--	--	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--
		Winter	--	v3.4	v3.4	v3.4	v3.3.1	--	--	v3.4	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--
		Spring	--	--	--	--	--	v3.3.1	v3.3.1	--	--	--	--	--	--	v3.3.1	v3.3.1	--
	12 UTC Initializations	Annual	--	v3.3.1	v3.3.1	--	v3.4	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.4	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Summer	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Fall	v3.3.1	--	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Winter	v3.3.1	v3.3.1	--	v3.4	v3.4	v3.4	v3.4	v3.3.1	v3.3.1	--	--	v3.4	--	--	v3.3.1	--
		Spring	v3.3.1	v3.3.1 *	v3.3.1 *	--	--	--	--	--	--	v3.3.1 *	v3.3.1 *	--	--	--	--	--

Table 6. SS (light shading) and PS (dark shading) pair-wise differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for surface wind BCRMSE and bias by season and forecast lead time for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Surface Wind Speed			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
BCRMSE	00 UTC Initializations	Annual	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	--	--	--	--
		Summer	--	--	--	v3.3.1	--	--	v3.4	--	--	--	--	--	--	--	--	--
		Fall	--	v3.3.1	v3.3.1	--	--	--	--	v3.3.1	v3.3.1	v3.3.1	--	--	v3.3.1	--	--	v3.3.1
		Winter	v3.3.1	v3.3.1	--	--	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	--	--	v3.3.1	--	--	--	--
		Spring	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	12 UTC Initializations	Annual	--	v3.3.1	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	--	v3.3.1	v3.3.1	--	v3.3.1
		Summer	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--
		Fall	v3.3.1	v3.3.1	v3.3.1	--	v3.3.1	--	--	--	v3.3.1	--	--	--	--	v3.3.1	--	--
		Winter	--	v3.3.1	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--	--	--	v3.3.1	--	--
		Spring	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Bias	00 UTC Initializations	Annual	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	v3.4	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	v3.4
		Summer	v3.4	v3.4	v3.4	v3.4	--	--	--	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	--	v3.4
		Fall	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	v3.4	v3.4	v3.4	--	--	--	--	--	--
		Winter	v3.4	--	--	v3.4	--	--	--	--	v3.4	v3.4	v3.4	v3.4	--	--	--	--
		Spring	v3.4	v3.4	--	--	--	--	--	--	v3.4	v3.4	v3.4	v3.4	--	--	--	--
	12 UTC Initializations	Annual	--	--	--	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	--	v3.4	v3.4	v3.4	v3.4	v3.4
		Summer	v3.3.1	--	--	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	--	--	v3.4	v3.4	v3.4	v3.4
		Fall	v3.4	v3.4	--	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	--	--	--	--	v3.4	--
		Winter	v3.4	--	--	v3.4	v3.4	v3.4	--	v3.4	--	--	--	--	--	--	--	--
		Spring	--	--	--	v3.4	v3.4	v3.4	v3.4	v3.4	--	--	--	--	v3.4	v3.4	--	v3.4

Table 7. SS differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for 3-hour QPF GSS by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

3-hour QPF			00 UTC Initializations									12 UTC Initializations								
			>0.01	>0.02	>0.05	>0.1	>0.15	>0.25	>0.35	>0.5	>1	>0.01	>0.02	>0.05	>0.1	>0.15	>0.25	>0.35	>0.5	>1
Gilbert Skill Score	Annual	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--
	Summer	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	v3.4	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	v3.3.1
	Fall	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	v3.3.1	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	v3.4
		f48	--	--	--	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--
	Winter	f12	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	v3.3.1	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	v3.4
		f36	--	--	--	--	--	--	v3.3.1	v3.3.1	--	--	--	--	--	--	--	--	--	--
		f48	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	Spring	f12	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f24	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f36	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
		f48	--	--	v3.3.1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	v3.4

Table 8. SS differences for the AFWA configuration run with WRF v3.3.1 and WRF v3.4 (where the highlighted configuration is favored) for 24-hour QPF GSS by season, forecast lead time, and threshold for the 00 UTC and 12 UTC initializations separately over the CONUS verification domain.

Daily QPF				>0.01	>0.25	>0.5	>0.75	>1	>1.25	>1.5	>2	>3
Gilbert Skill Score	00 UTC Initializations	Annual	f36	v3.3.1	v3.3.1	--	--	--	--	--	--	v3.3.1
		Summer	f36	--	--	--	--	--	--	--	--	--
		Fall	f36	v3.3.1	--	--	--	--	--	--	--	--
		Winter	f36	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1	v3.3.1	--	--	--
		Spring	f36	v3.3.1	--	--	--	--	--	--	--	v3.3.1
	12 UTC Initializations	Annual	f24	v3.3.1	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Summer	f24	--	--	--	--	--	--	--	v3.4	--
			f48	--	--	--	--	--	v3.3.1	v3.3.1	v3.3.1	v3.3.1
		Fall	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	--	--	--	--
		Winter	f24	v3.3.1	--	--	--	--	--	--	--	v3.4
			f48	v3.3.1	--	--	--	--	--	--	--	--
		Spring	f24	--	--	--	--	--	--	--	--	--
			f48	--	--	--	--	--	v3.4	v3.4	--	--

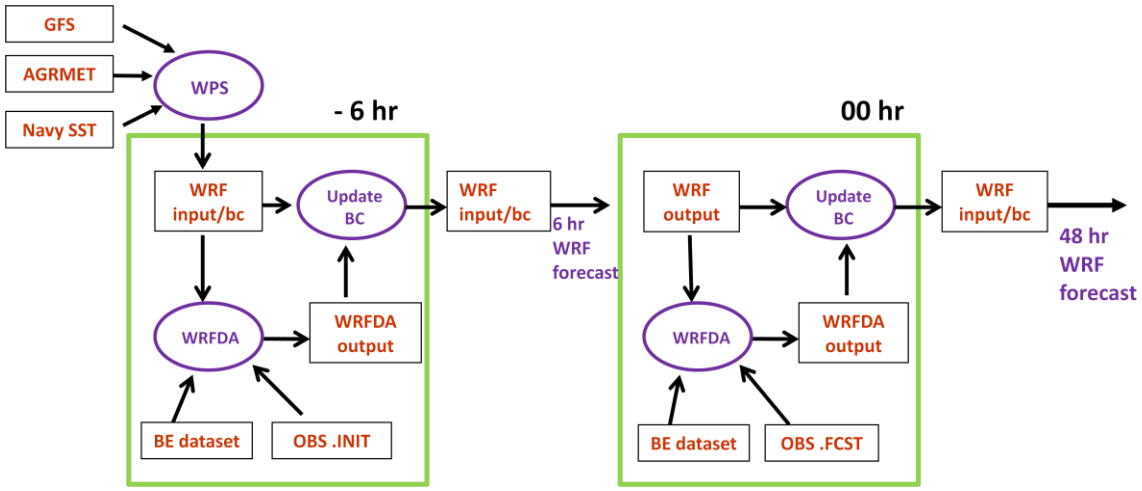


Figure 1. Overview of 6-hr “warm start” spin-up.

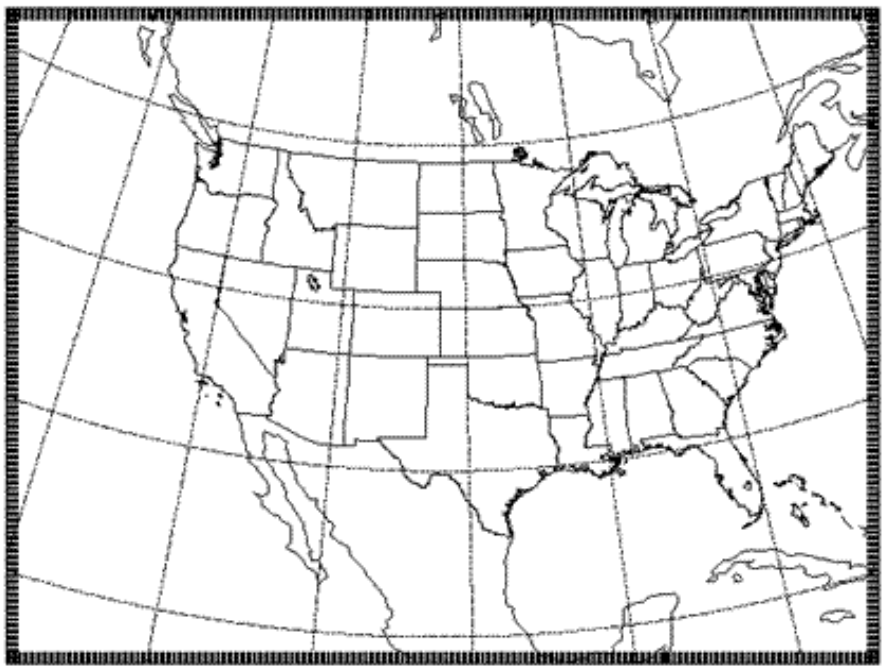


Figure 2. WRF-ARW computational domain.

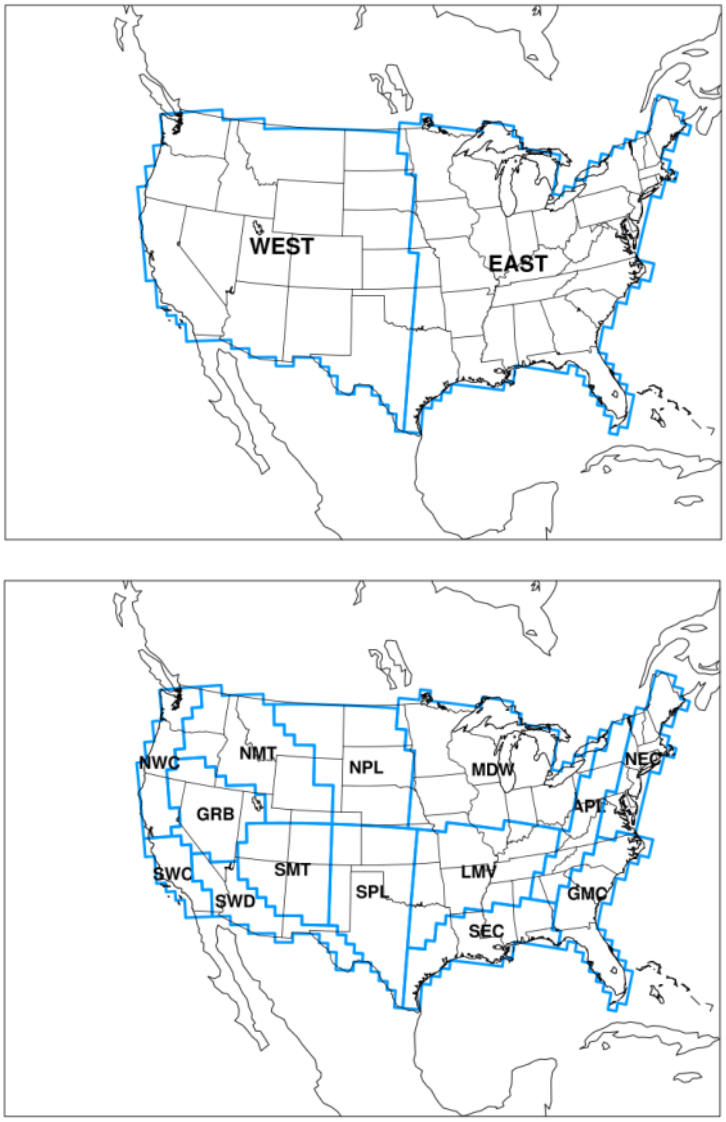
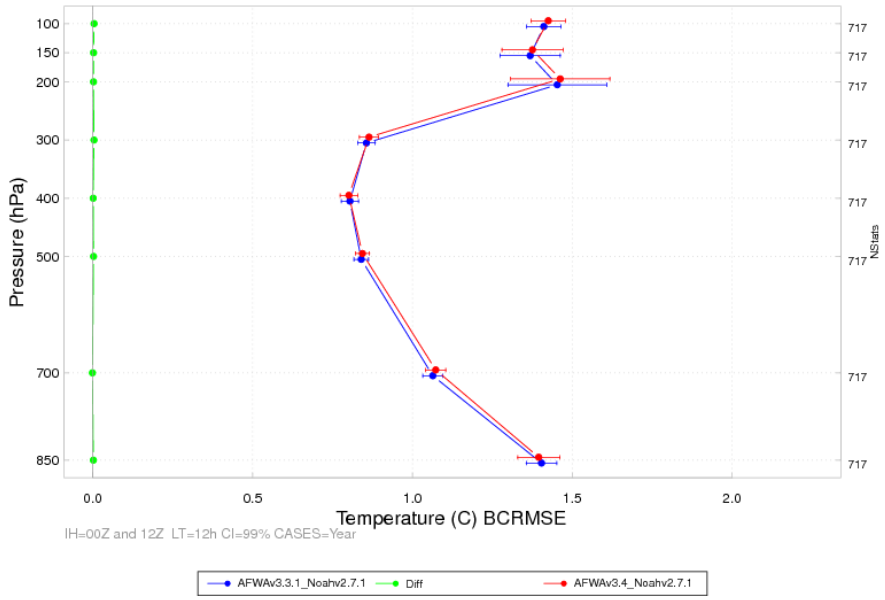


Figure 3. Map showing the locations of the CONUS-West, CONUS-East (top) and 14 regional verification domains (bottom). The outermost outline of the regional domains depict the CONUS verification domain.

(a) LT=12 h



(b) LT=48 h

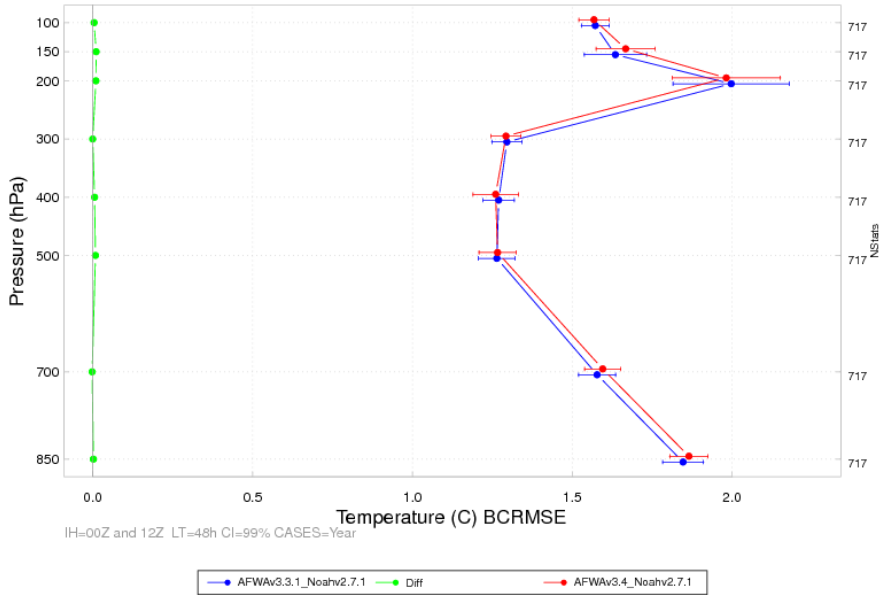
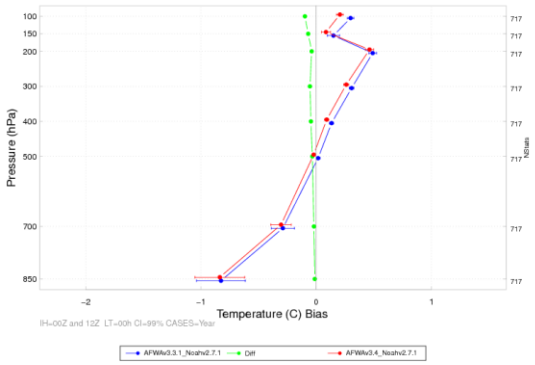
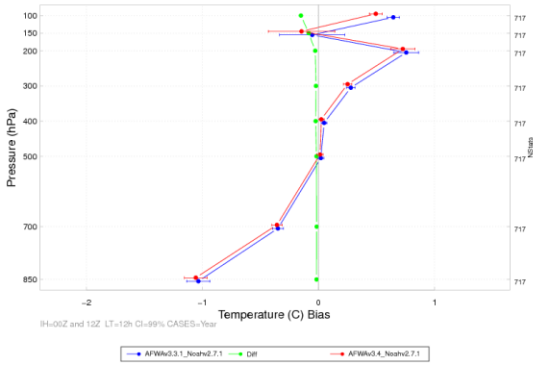


Figure 4. Vertical profile of the median BCRMSE for temperature (°C) for the CONUS verification domain aggregated across the entire year of cases for the (a) 12- and (b) 48-h lead times. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

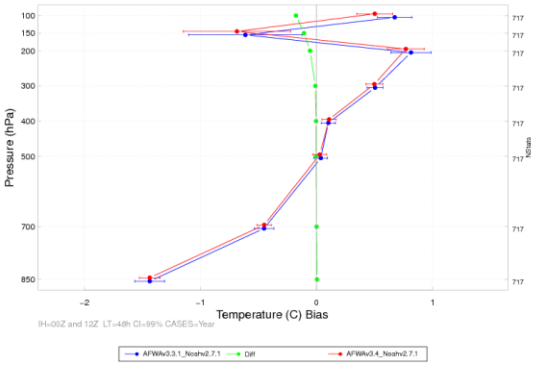
(a) Annual LT=0 h



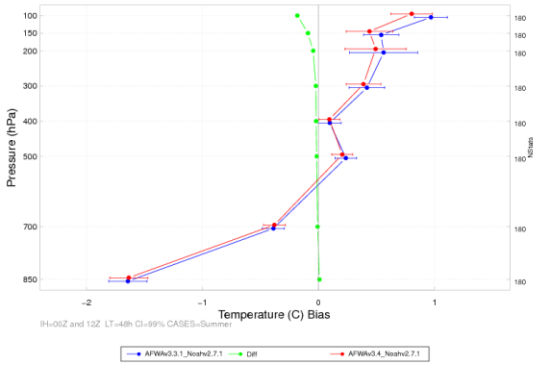
(b) Annual LT=12 h



(c) Annual LT=48 h



(d) Summer LT=48 h



(e) Winter LT=48 h

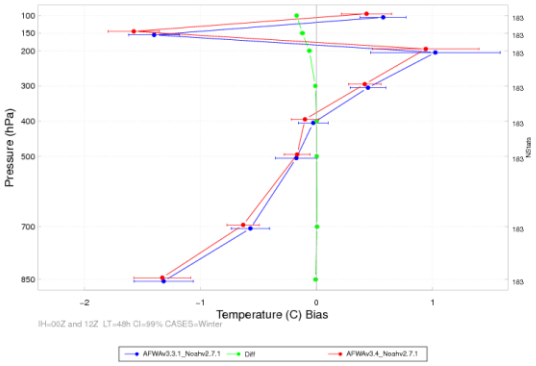
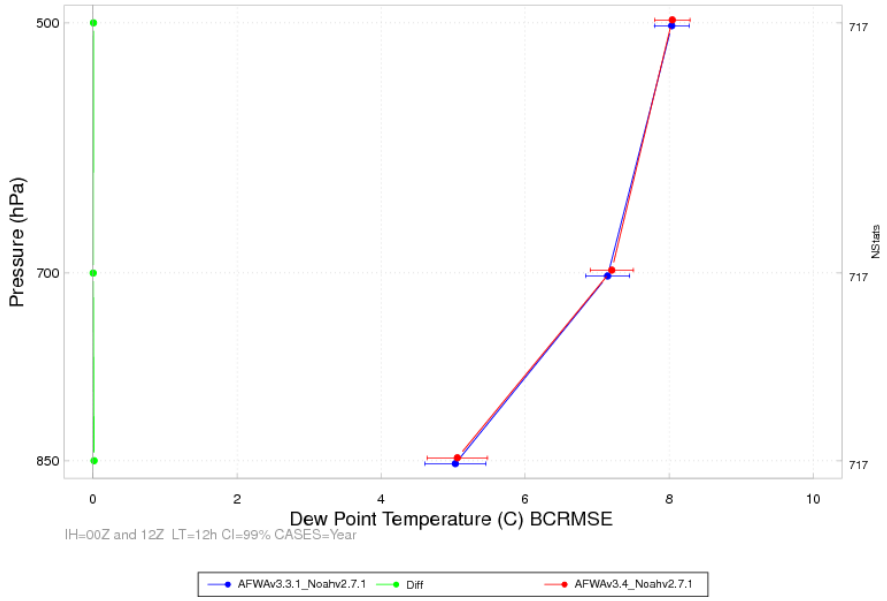


Figure 5. Vertical profile of the median bias for temperature (°C) for the CONUS verification domain aggregated across the entire year of cases for the (a) 0-, (b) 12- and (c) 48-h lead times and for 48-h lead time for the (d) summer aggregation and (e) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

(a) LT=12 h



(b) LT=48 h

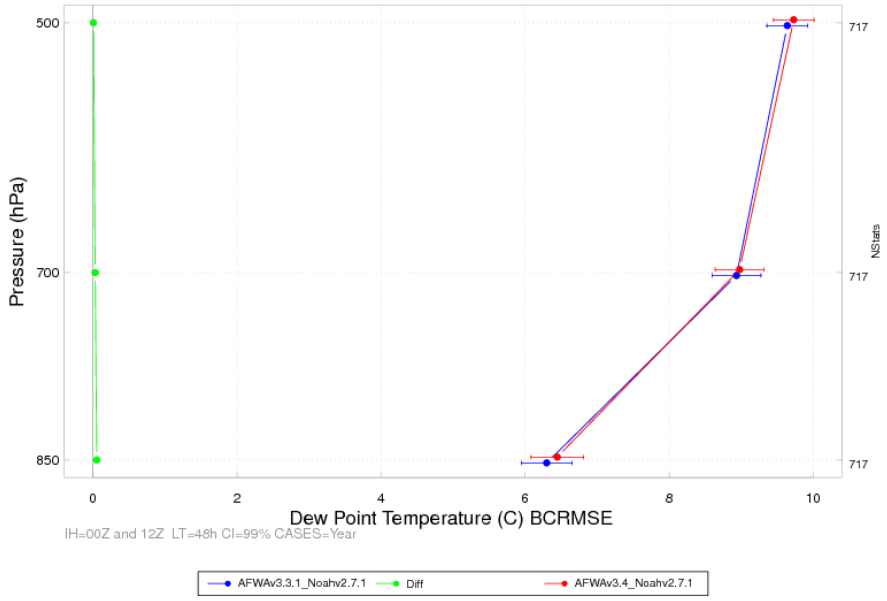
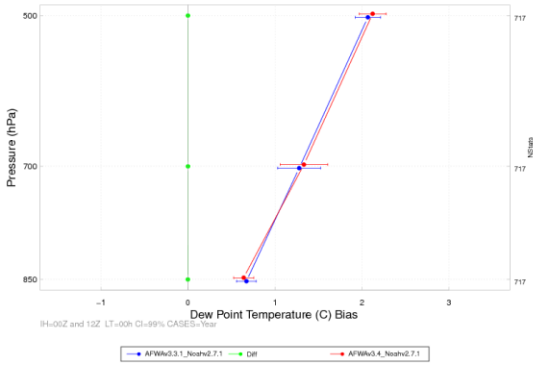
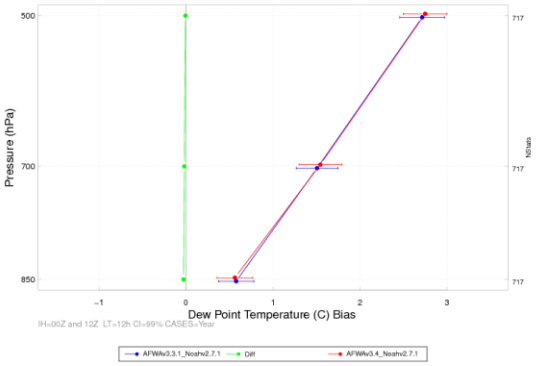


Figure 6. Vertical profile of the median BCRMSE for dew point temperature (°C) for the CONUS verification domain aggregated across the entire year of cases for the (a) 12- and (b) 48-h lead times. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

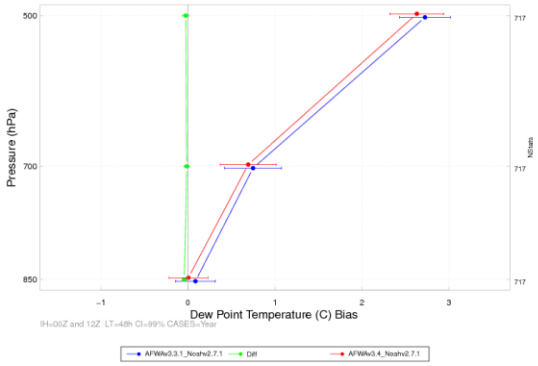
(a) Annual LT=0 h



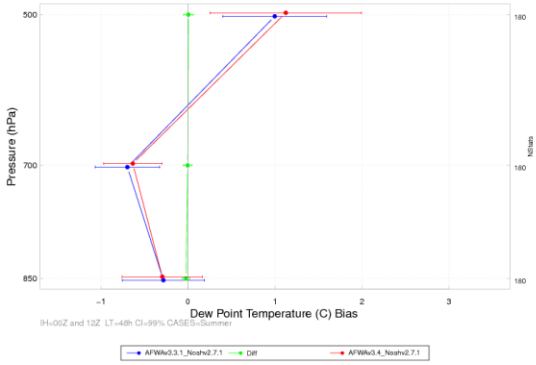
(b) Annual LT=12 h



(c) Annual LT=48 h



(d) Summer LT=48 h



(e) Winter LT=48 h

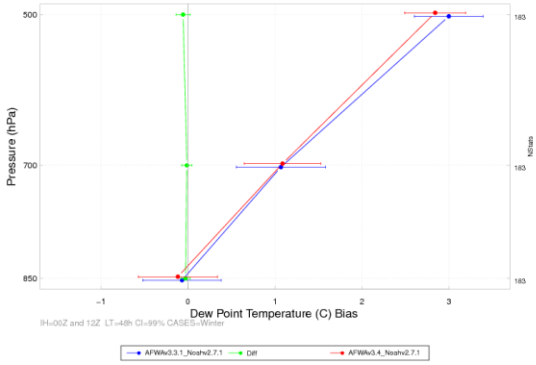
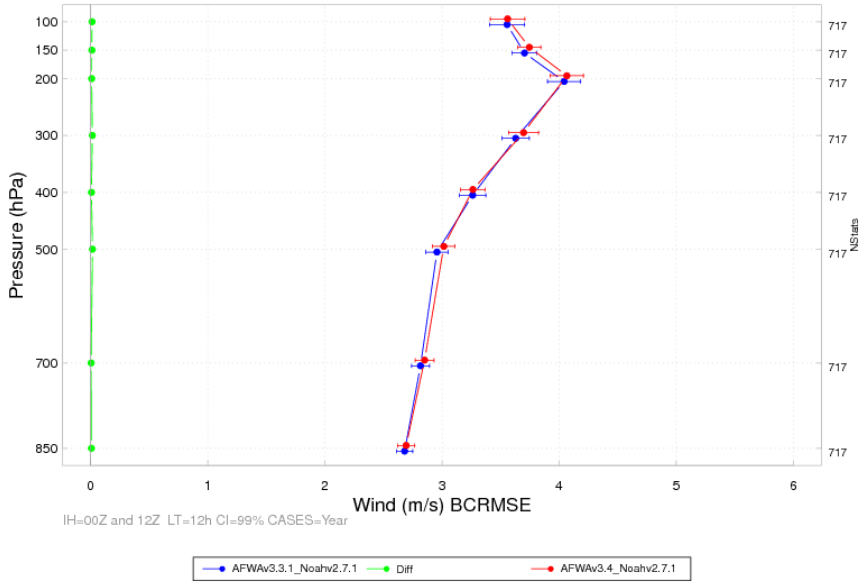


Figure 7. Vertical profile of the median bias for dew point temperature (°C) for the CONUS verification domain aggregated across the entire year of cases for the (a) 0-, (b) 12- and (c) 48-h lead times and for 48-h lead time for the (d) summer aggregation and (e) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

(a) LT=12 h



(b) LT=48 h

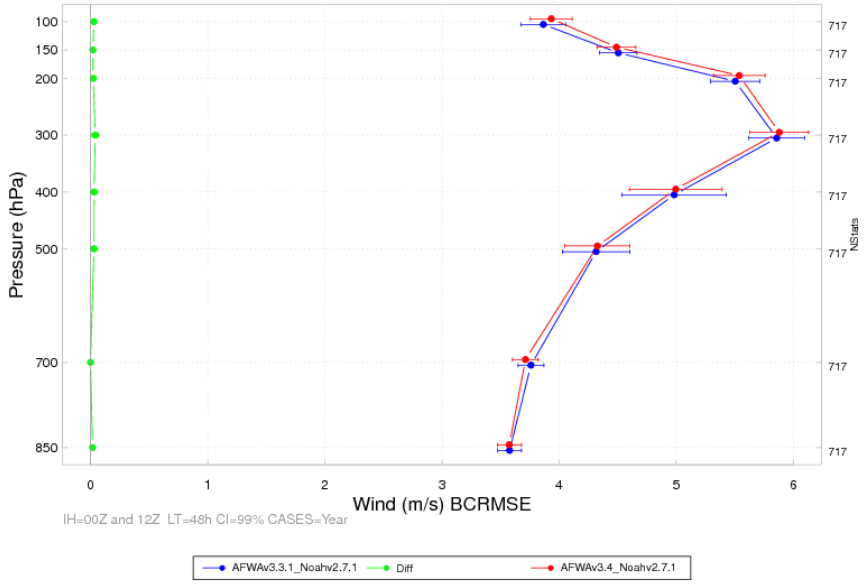


Figure 8. Vertical profile of the median BCRMSE for wind speed (m s^{-1}) for the CONUS verification domain aggregated across the entire year of cases for the (a) 12- and (b) 48-h lead times. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

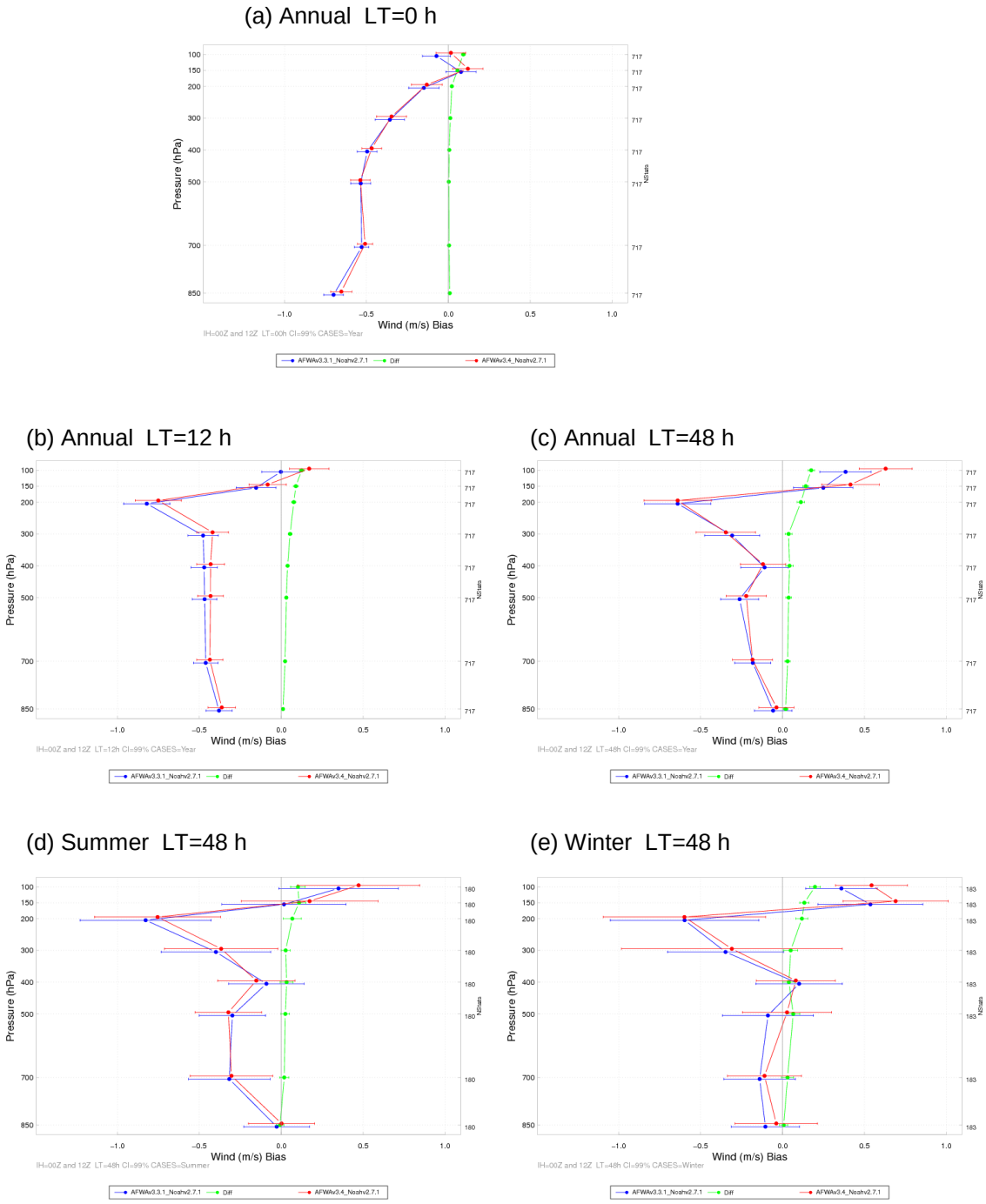
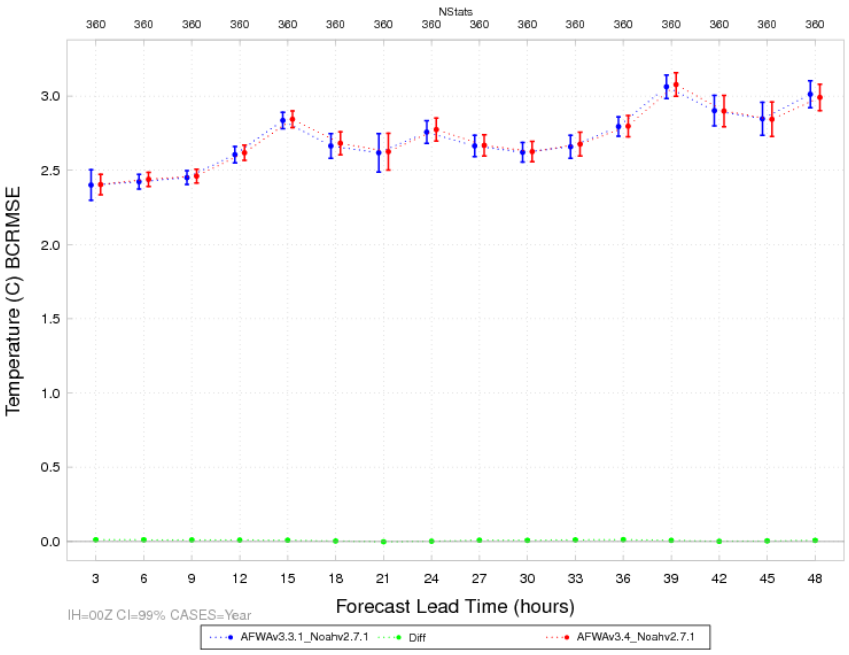


Figure 9. Vertical profile of the median bias for wind speed (m s^{-1}) for the CONUS verification domain aggregated across the entire year of cases for the (a) 0-, (b) 12- and (c) 48-h lead times and for 48-h lead time for the (d) summer aggregation and (e) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

(a) Annual IH=00 UTC



(b) Annual IH=12 UTC

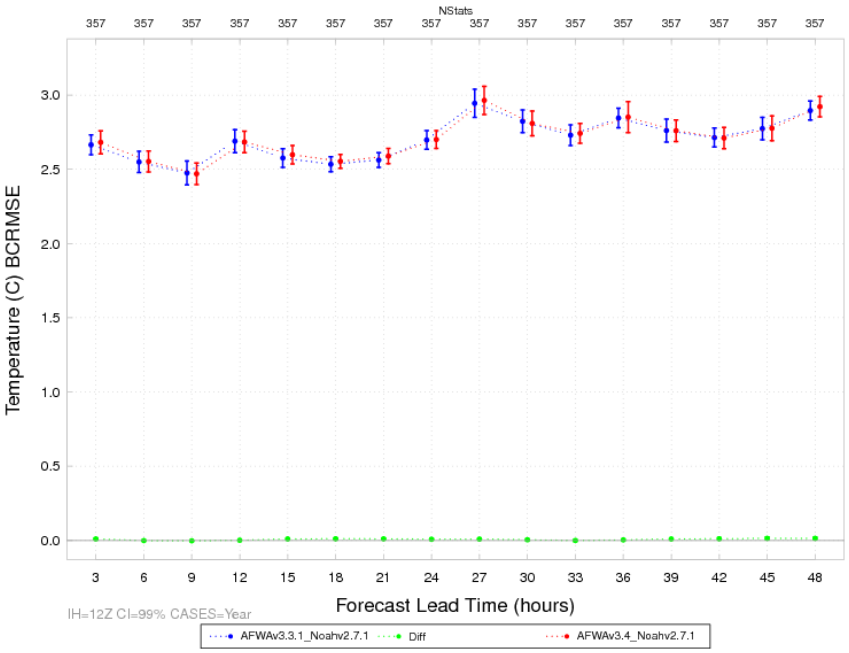
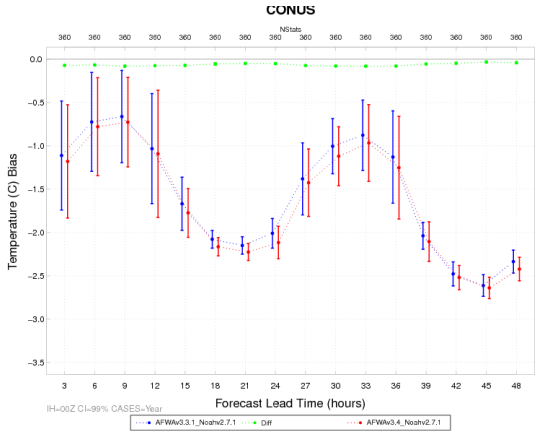
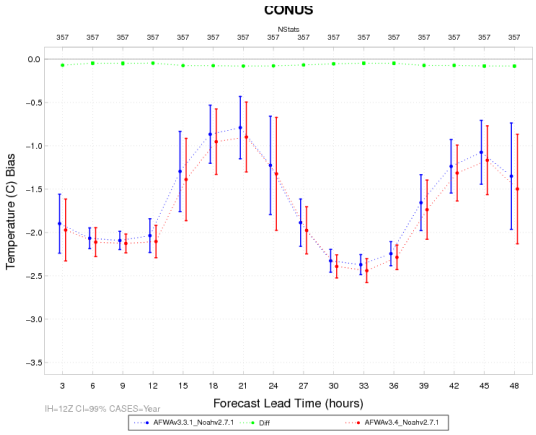


Figure 10. Time series plot of 2 m AGL temperature (°C) for median BCRMSE for the (a) 00 UTC initializations and (b) 12 UTC initializations aggregated across the entire year of cases. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

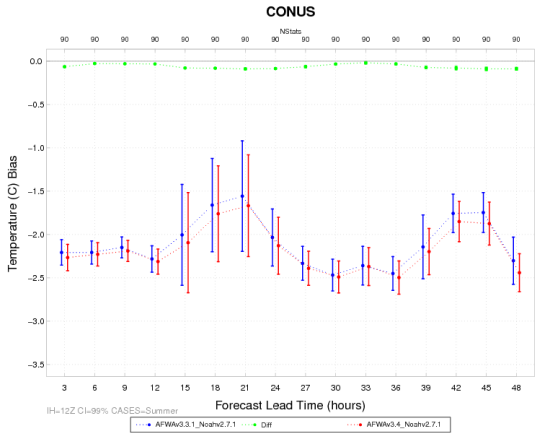
(a) Annual IH=00 UTC



(b) Annual IH=12 UTC



(c) Summer IH=12 UTC



(d) Winter IH=12 UTC

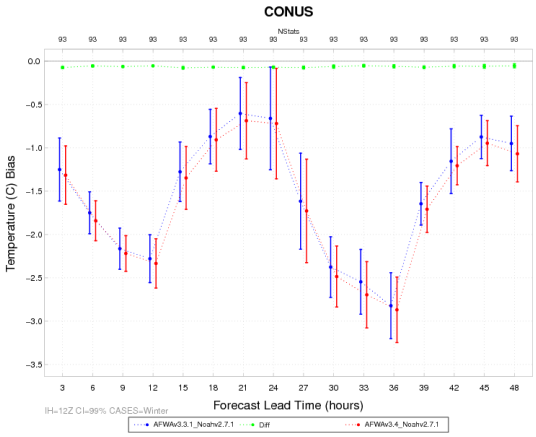


Figure 11. Time series plot of 2 m AGL temperature (°C) for median bias for the CONUS verification domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 12 UTC initializations for the (c) summer aggregation and (d) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

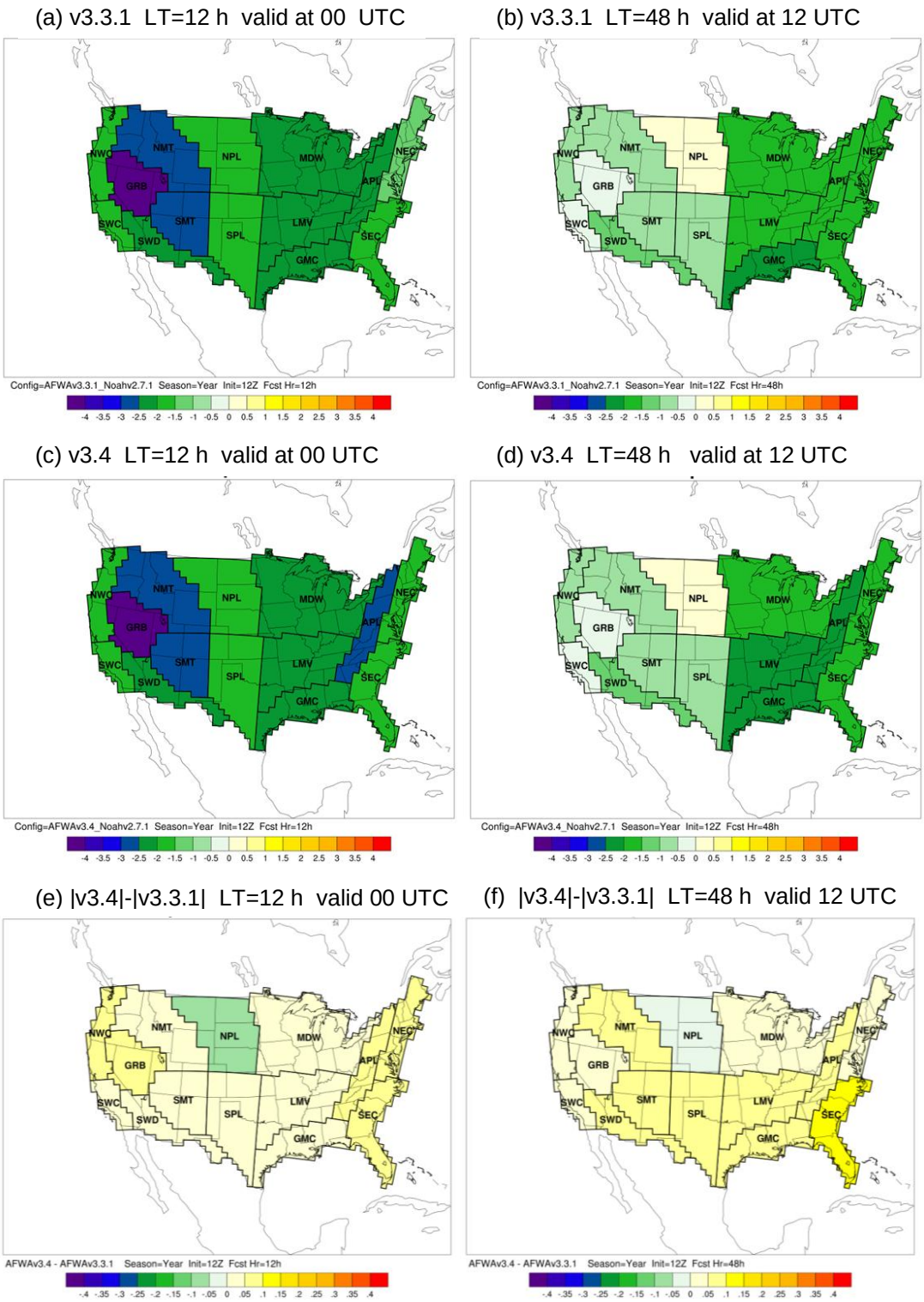
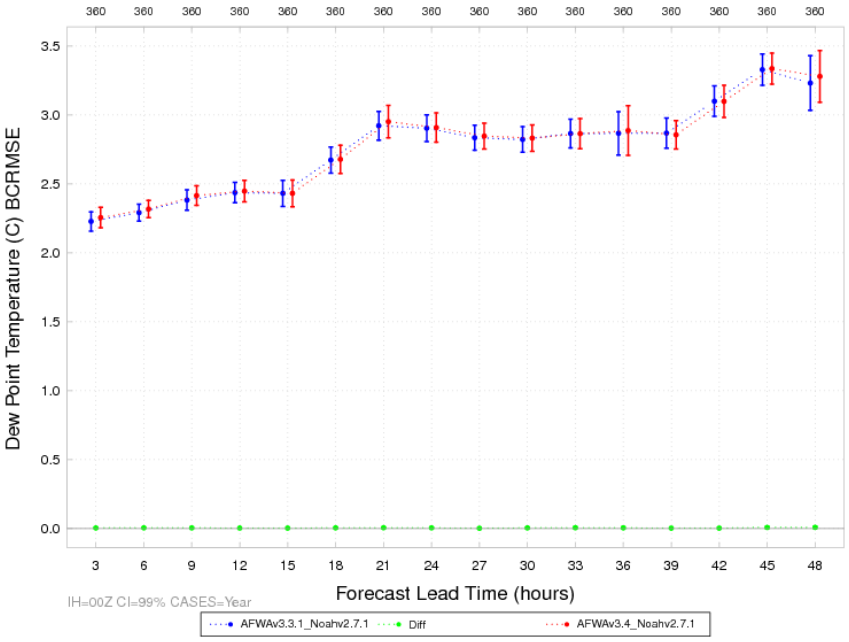


Figure 12. Regional median bias in 2 m AGL temperature (°C) for the 12 UTC initializations, and (a) the WRFv3.3.1 configuration at 12 h lead time (b) WRFv3.3.1 configuration at 48 h lead time (c) the WRFv3.4 configuration at 12 h lead time (d) the WRFv3.4 configuration at 48 h lead time (e) the magnitude (absolute value) difference between WRFv3.4 and WRFv3.3.1 at 12 h lead time (f) the magnitude difference between WRFv3.4 and WRFv3.3.1 at 48 h lead time, aggregated across the entire year of cases.

(a) IH=00 UTC



(b) IH=12 UTC

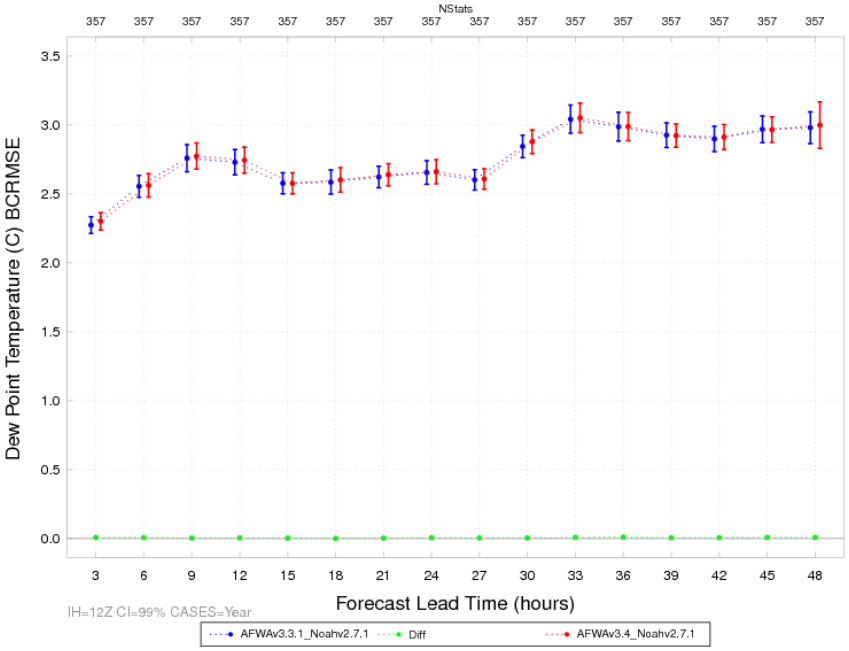


Figure 13. Time series plot of 2 m AGL dew point temperature (°C) for median BCRMSE for the (a) 00 UTC initializations and (b) 12 UTC initializations aggregated across the entire year of cases. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

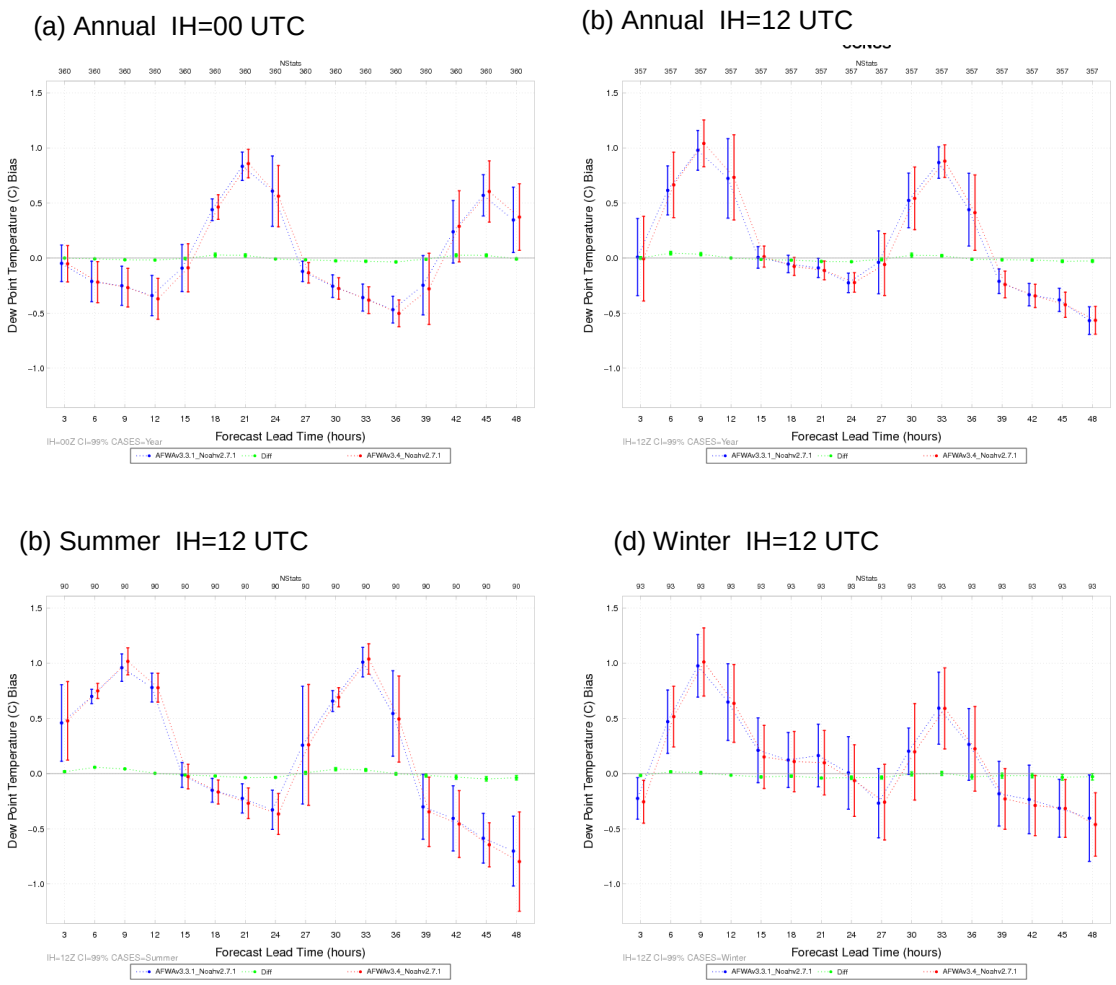
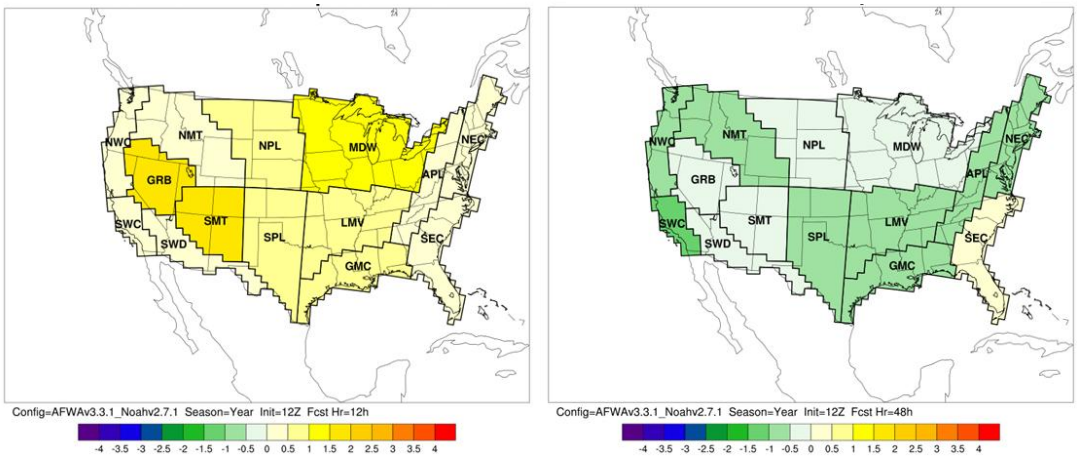


Figure 14. Time series plot of 2 m AGL dew point temperature (°C) for median bias for the CONUS verification domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 12 UTC initializations for the (c) summer aggregation and (d) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

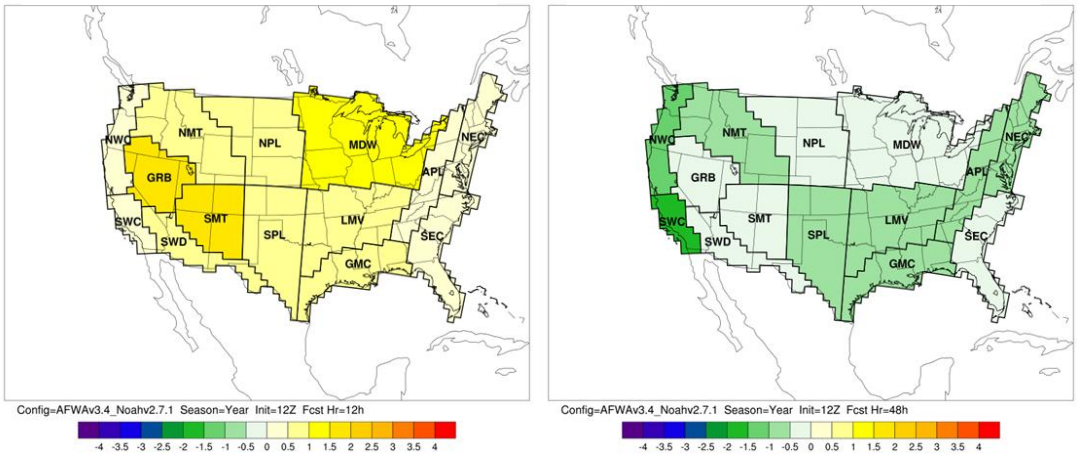
(a) v3.3.1 LT=12 h valid at 00 UTC

(b) v3.3.1 LT=48 h valid at 12 UTC



(c) v3.4 LT=12 h valid at 00 UTC

(d) v3.4 LT=48 h valid at 12 UTC



(e) |v3.4|-|v3.3.1| LT=12 h valid 00 UTC

(f) |v3.4|-|v3.3.1| LT=48 h valid 12 UTC

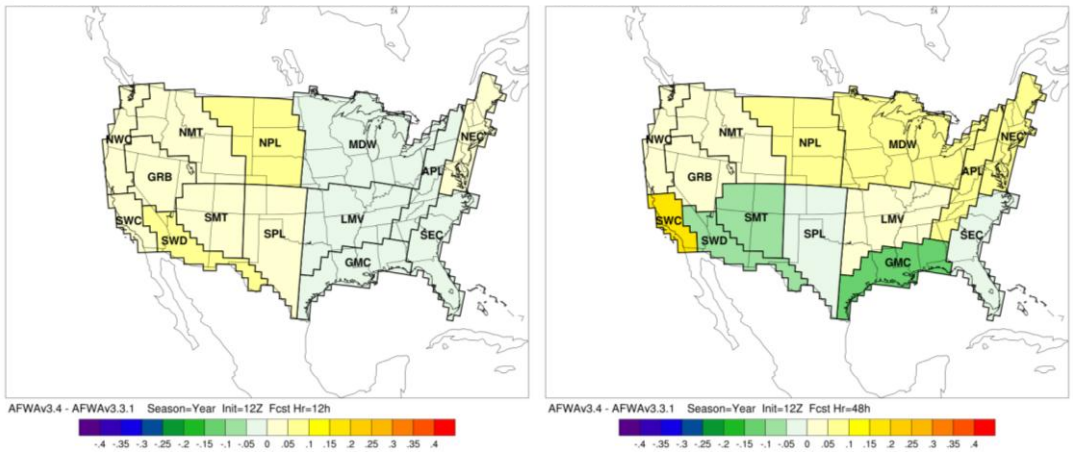
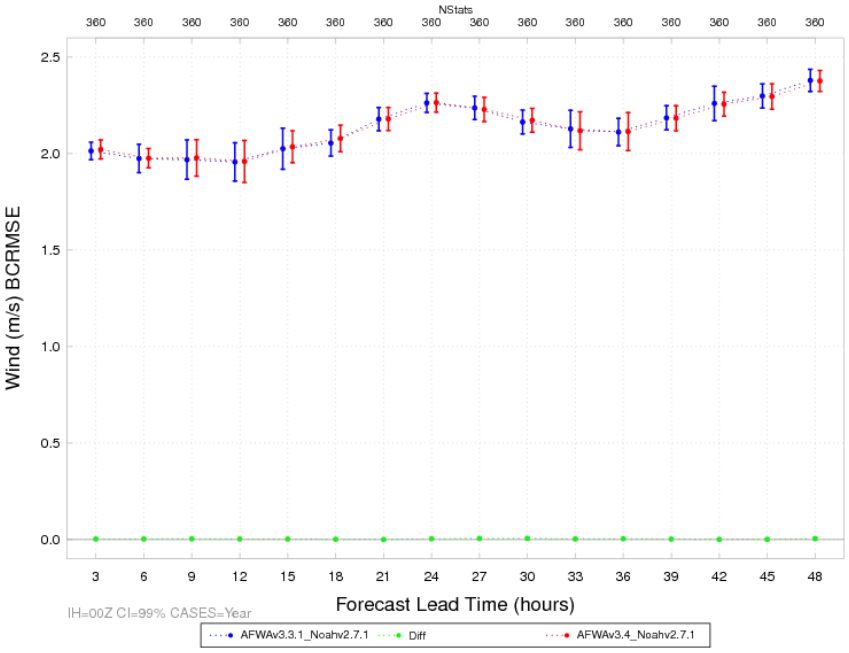


Figure 15. Regional median bias in 2 m AGL dew point temperature (°C) for the 12 UTC initializations, and (a) the WRFv3.3.1 configuration at 12 h lead time (b) WRFv3.3.1 configuration at 48 h lead time (c) the WRFv3.4 configuration at 12 h lead time (d) the WRFv3.4 configuration at 48 h lead time (e) the magnitude (absolute value) difference between WRFv3.4 and WRFv3.3.1 at 12 h lead time (f) the magnitude difference between WRFv3.4 and WRFv3.3.1 at 48 h lead time, aggregated across the entire year of cases.

(a) IH=00 UTC



(b) IH=12 UTC

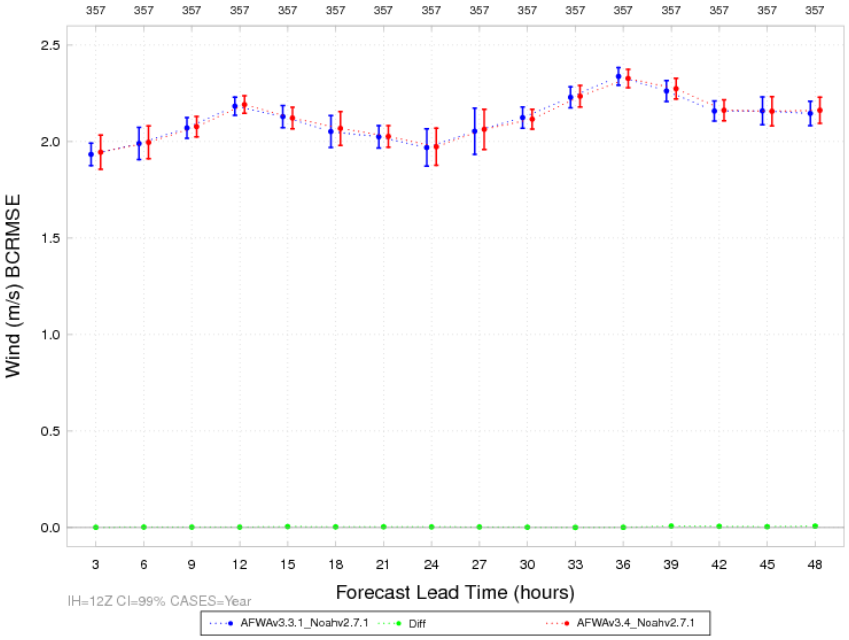


Figure 16. Time series plot of 10 m AGL wind speed (m s^{-1}) for median BCRMSE for the 12 UTC initializations (a) aggregated across the entire year of cases and (b) aggregated across the summer season. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

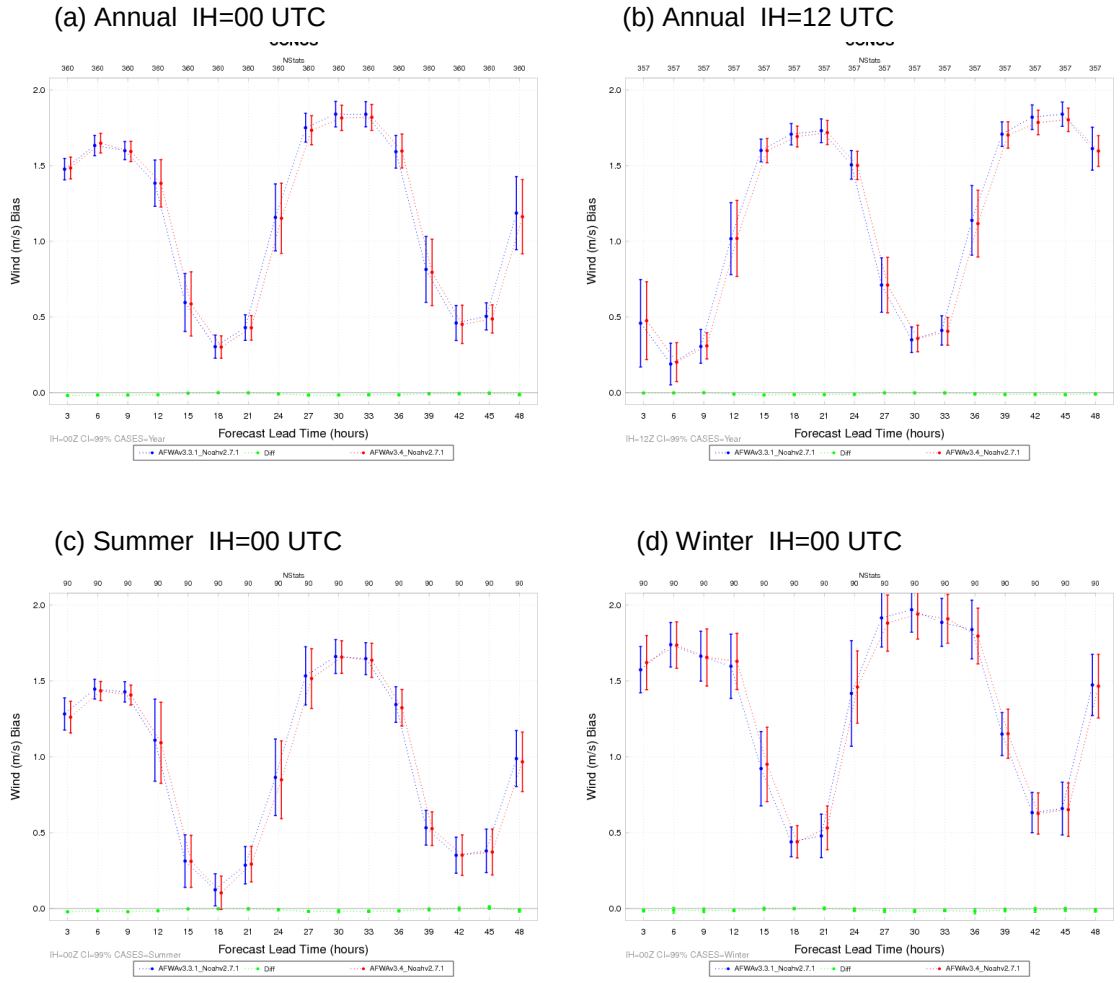


Figure 17. Time series plot of 10 m AGL wind speed (m s^{-1}) for median bias for the CONUS verification domain aggregated across the entire year of cases for the (a) 00 UTC initializations and (b) 12 UTC initializations and for the 00 UTC initializations for the (c) summer aggregation and (d) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

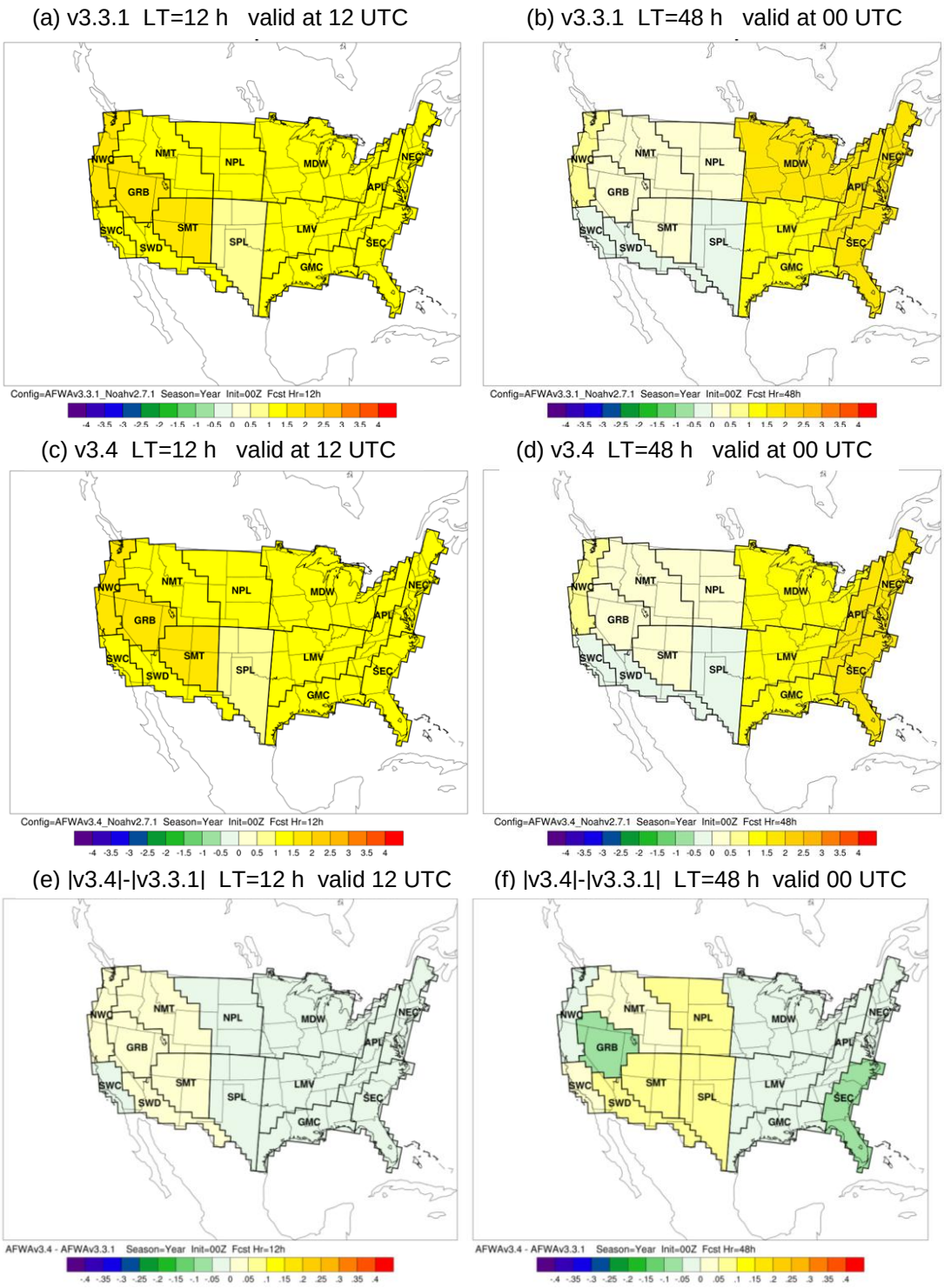


Figure 18. Regional median bias in 10 m AGL wind speed (m s^{-1}) for the 00 UTC initialization, and (a) the WRFv3.3.1 configuration at 12 h lead time (b) WRFv3.3.1 configuration at 48 h lead time (c) the WRFv3.4 configuration at 12 h lead time (d) the WRFv3.4 configuration at 48 h lead time (e) the magnitude (absolute value) difference between WRFv3.4 and WRFv3.3.1 at 12 h lead time (f) the magnitude difference between WRFv3.4 and WRFv3.3.1 at 48 h lead time, aggregated across the entire year of cases.

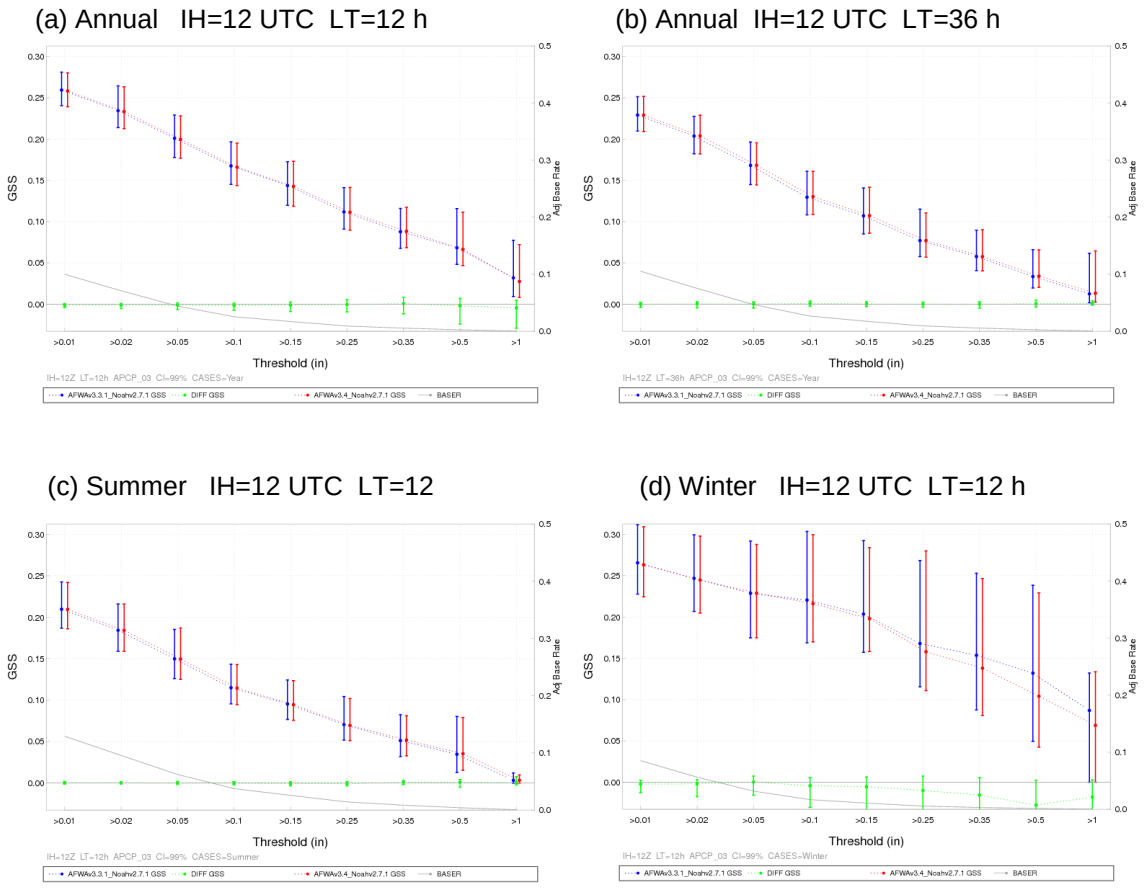


Figure 19. Threshold series plot of 3-h accumulated precipitation (in) for median GSS for the 12 UTC initializations aggregated across the entire year of cases for the (a) 12-h lead time and the (b) 36-h lead time and for the 12 UTC initializations for the 12-h lead time for the (c) summer aggregation and (d) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

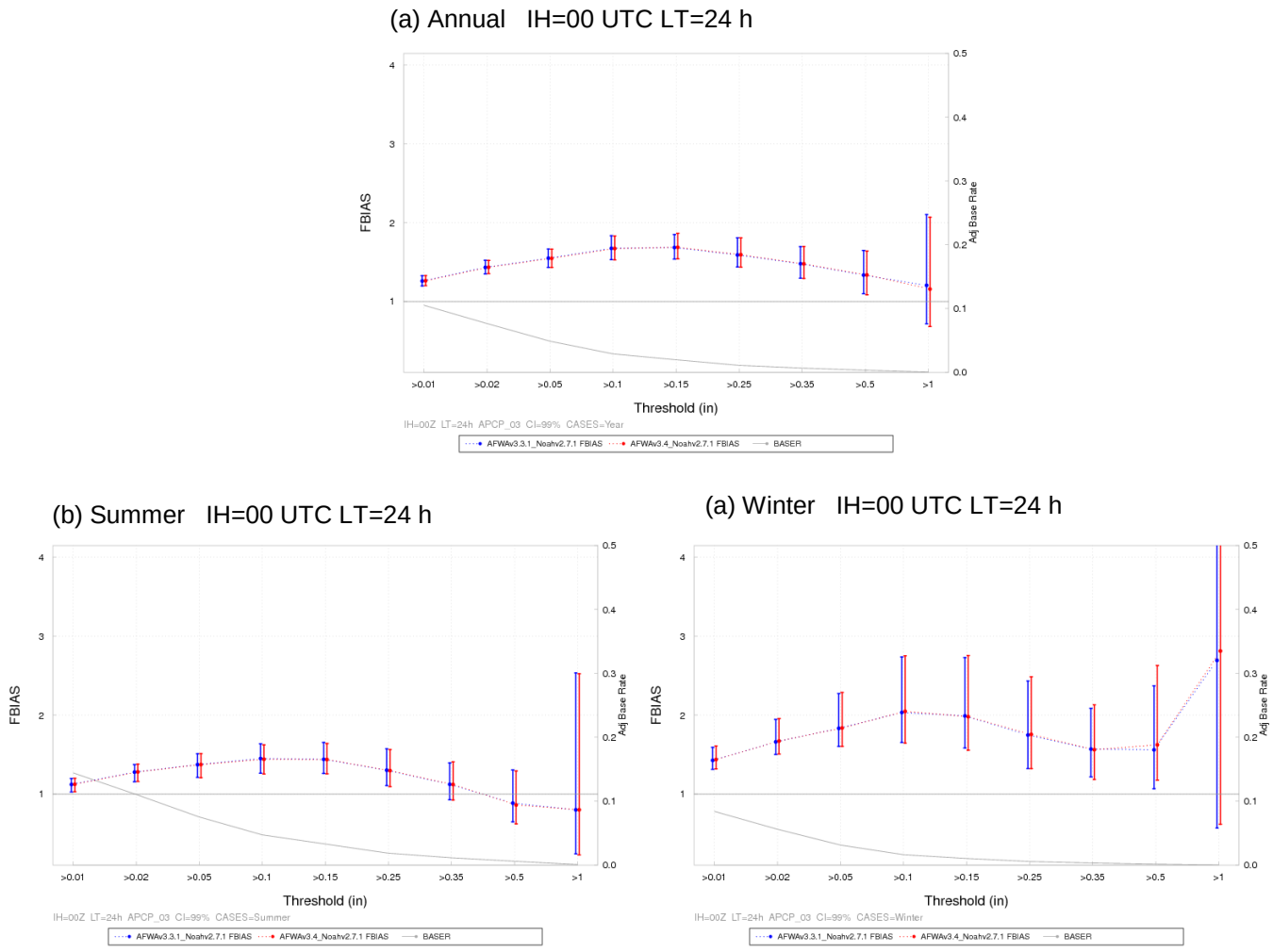
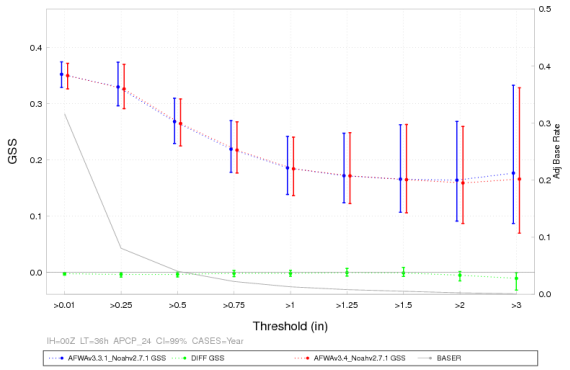
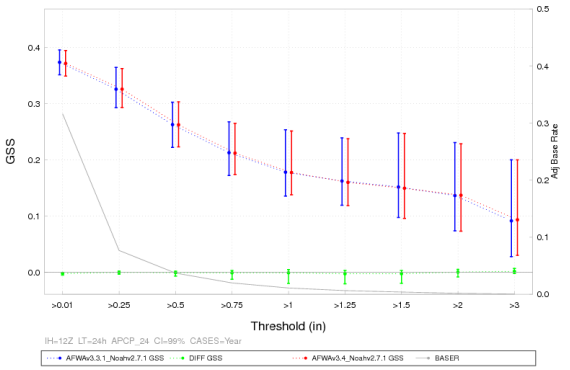


Figure 20. Threshold series plot of 3-h accumulated precipitation (in) for median frequency bias for the 00 UTC initialization for the 24-h lead time aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

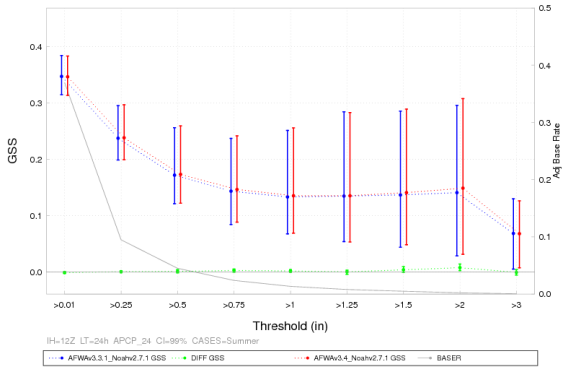
(a) Annual IH=00 UTC LT=36 h



(b) Annual IH=12 UTC LT=24 h



(c) Summer IH=12 UTC LT=24 h



(d) Winter IH=12 UTC LT=24 h

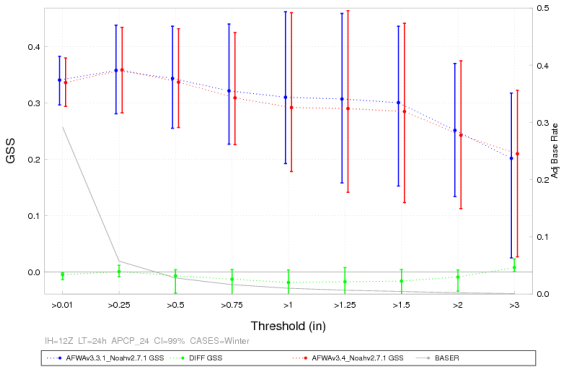
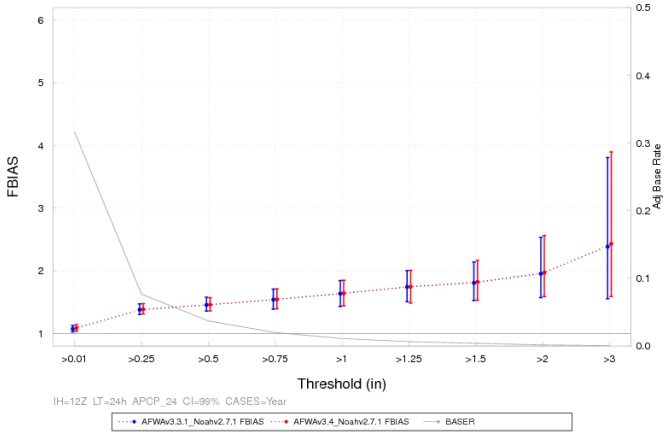
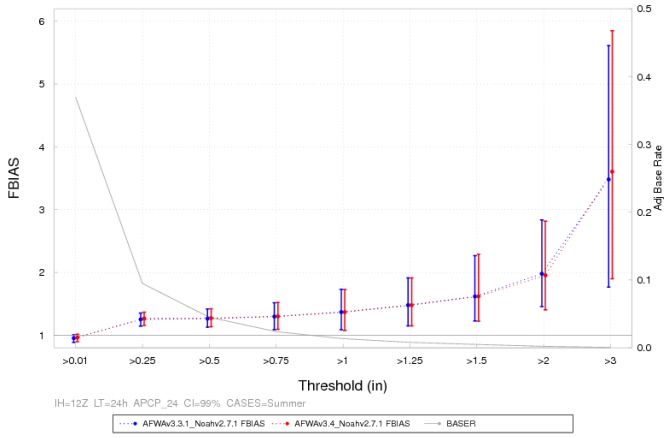


Figure 21. Threshold series plot of 24-h accumulated precipitation (in) for median GSS for the (a) 00 UTC for the 36-h lead time aggregated across the entire year of cases, the 12 UTC initialization for the 24-h lead time aggregated across the (b) entire year of cases, (c) summer aggregation, and (d) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

(a) Annual IH=12 UTC LT=24 h



(b) Summer IH=12 UTC LT=24 h



(c) Winter IH=12 UTC LT=24 h

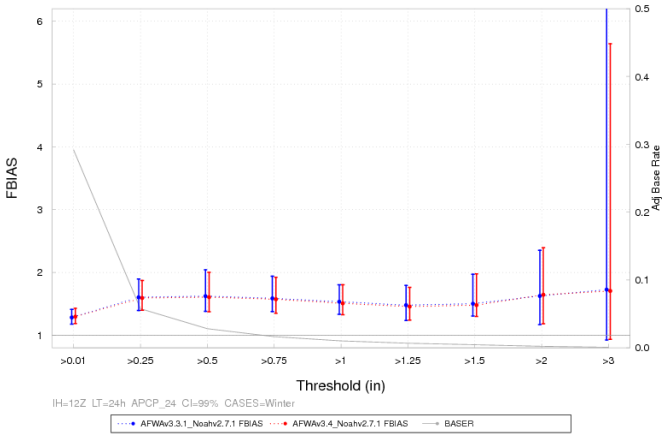


Figure 22. Threshold series plot of 24-h accumulated precipitation (in) for median frequency bias for the 12 UTC initializations for the 24-h lead time aggregated across the (a) entire year of cases, (b) summer aggregation, and (c) winter aggregation. The WRFv3.3.1 configuration is in blue, the WRFv3.4 configuration in red, and the pair-wise differences (WRFv3.4 – WRFv3.3.1) in green. The horizontal bars attached to the median represent the 99% CIs.

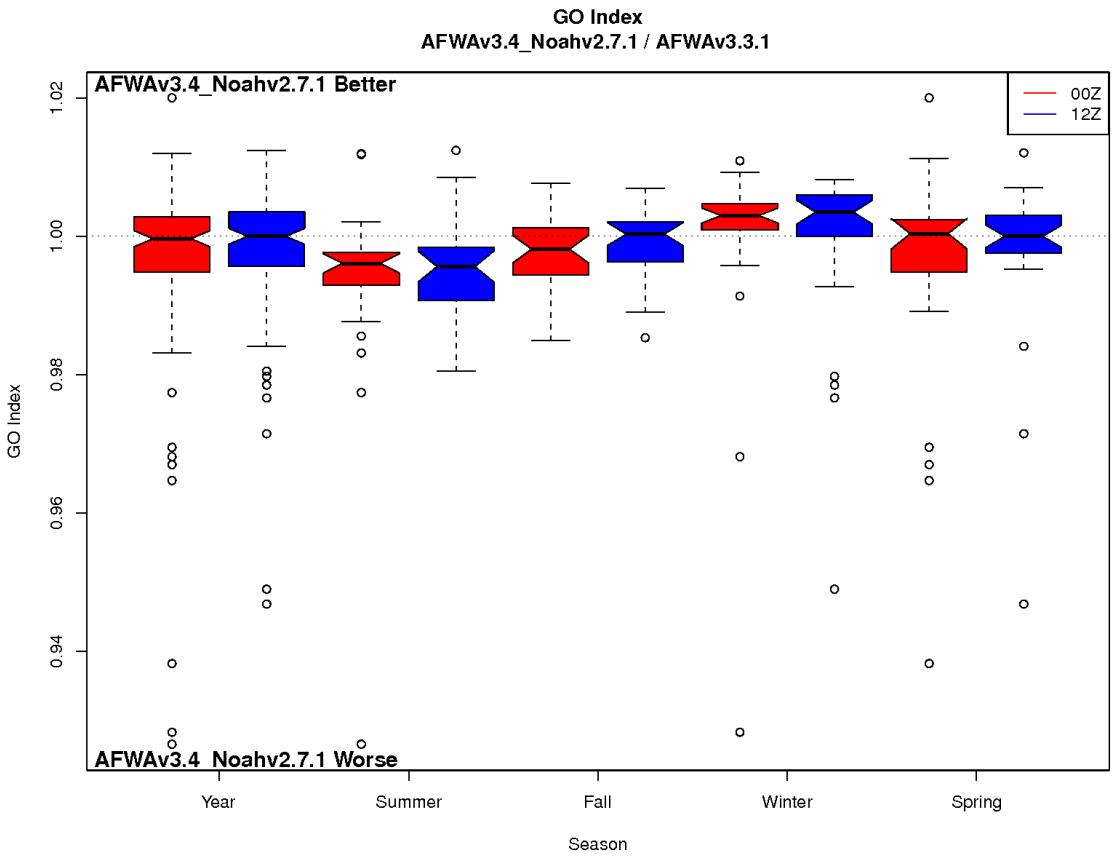


Figure 23. Boxplot of GO Index values aggregated across the entire year of cases and for all seasons, stratified by initialization time where 00 UTC is in red and 12 UTC is in blue. The median value is the thick black line located at the vertex of the notches, the notches around the median is an approximation of the 95% confidence about the median, the whiskers, denoted by the black, dashed lines, denote the largest values that are not outliers, and the circles represent the outliers.

Appendix A: Case list. Dates in bold were not included in the verification due to bad or missing input data.

00 UTC Initialization	12 UTC Initialization
July 2011: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31	July 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
August 2011: 3, 6, 9, 12, 15, 18, 21, **24**, 27, 30	August 2011: **1**, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
September 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	September 2011: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
October 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	October 2011: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
November 2011: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28	November 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
December 2011: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31	December 2011: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29
January 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	January 2012: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
February 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	February 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27
March 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30	March 2012: 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31
April 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	April 2012: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30
May 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29	May 2012: **3**, **6**, 9, 12, 15, 18, 21, 24, 27, 30
June 2012: 1, **4**, 7, 10, 13, 16, 19, 22, 25, 28	June 2012: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29

Appendix B: Subset a WRF ***namelist.input*** used in this test

```
&wrfva4
thin_conv       = .TRUE.,
use_synopobs    = .TRUE.,
use_shipsobs    = .TRUE.,
use_metarobs    = .TRUE.,
use_soundobs    = .TRUE.,
use_mtgirsobs   = .TRUE.,
use_tamdarobs   = .TRUE.,
use_pilotobs    = .TRUE.,
use_airepobs    = .TRUE.,
use_geoamvobs   = .TRUE.,
use_polaramvobs  = .TRUE.,
use_buoyobs     = .TRUE.,
use_profilerobs   = .TRUE.,
use_satemobs    = .TRUE.,
use_gpspwobs    = .TRUE.,
use_gpsrefobs   = .TRUE.,
top_km_gpsro    = 30.0,
bot_km_gpsro    = 0.0,
use_ssmiretrievalobs = .TRUE.,
use_qscatobs    = .TRUE.,

&wrfvar6
max_ext_its     = 2,
ntmax           = 200,
nsave           = 4,
write_interval = 5,
```

```
eps           = 1.E-02,
precondition_cg = .FALSE.,
precondition_factor = 1.0,
use_lanczos    = .FALSE.,
orthonorm_gradient = .FALSE.,

&time_control
run_hours                           = 48,
interval_seconds                    = 10800,
history_interval                    = 180,
frames_per_outfile                  = 1,
restart                             = .false.,
io_form_history                     = 2,
input_outname                       = "wrfinput_d<domain>_<date>",
/

&domains
time_step                           = 90,
time_step_fract_num                 = 0,
time_step_fract_den                 = 1,
max_dom                             = 1,
e_we                                = 403,
e_sn                                = 302,
e_vert                              = 57,
num_metgrid_levels                  = 27,
num_metgrid_soil_levels             = 4,
dx                                  = 15000,
dy                                  = 15000,
p_top_requested                     = 1000,
interp_type                         = 1,
lowest_lvl_from_sfc                 = .false.,
lagrange_order                      = 1,
force_sfc_in_vinterp                = 6,
zap_close_levels                    = 500,
adjust_heights                      = .false.,
eta_levels                          = 1.000, 0.997, 0.992, 0.985, 0.978, 0.969, 0.960, 0.950,
                                      0.938, 0.925, 0.910, 0.894, 0.876, 0.857, 0.835, 0.812,
                                      0.787, 0.760, 0.731, 0.700, 0.668, 0.635, 0.600, 0.565,
                                      0.530, 0.494, 0.458, 0.423, 0.388, 0.355, 0.323, 0.293,
                                      0.264, 0.237, 0.212, 0.188, 0.167, 0.147, 0.130, 0.114,
                                      0.099, 0.086, 0.074, 0.064, 0.054, 0.046, 0.039, 0.032,
                                      0.027, 0.022, 0.017, 0.013, 0.010, 0.007, 0.004, 0.002,
                                      0.000,
/

&physics
mp_physcis                          = 4,
ra_lw_physics                       = 1,
ra_sw_physics                       = 1,
radt                                = 30,
sf_sfclay_physics                   = 1,
sf_surface_physics                  = 2,
bl_pbl_physics                      = 1,
bldt                                = 0,
cu_physics                          = 1,
cudt                                = 5,
```

surface_input_source = 1,
num_soil_layers = 4,
num_land_cat = 28,
mp_zero_out = 2,
/

&dynamics
rk_ord = 3,
diff_6th_opt = 2,
diff_6th_factor = 0.10
w_damping = 1,
diff_opt = 1,
km_opt = 4,
damp_opt = 3,
zdamp = 5000.,
dampcoef = 0.05
khdif = 0,
kvdif = 0,
smdiv = 0.1,
emdiv = 0.01,
epssm = 0.1,
time_step_sound = 0,
h_mom_adv_order = 5,
v_mom_adv_order = 3,
h_sca_adv_order = 5,
v_sca_adv_order = 3,
moist_adv_opt = 1,
scalar_adv_opt = 0,
chem._adv_opt = 0,
tke_adv_opt = 0,
/

&bdy_control
spec_bdy_width = 5,
spec_zone = 1,
relax_zone = 4,
specified = .true.,
/